Development of RNA aptamers and synthetic riboswitch using Capture-SELEX

Dem Fachbereich Biologie der Technischen Universität Darmstadt

zur

Erlangung des akademischen Grades

eines Doctor rerum naturalium

vorgelegte Dissertation von

M.Sc. Adrien Boussebayle

aus Dax

1. Referentin: Prof. Dr. Beatrix Süß

2. Referent: Prof. Dr. Alexander Loewer

Tag der Einreichung: 15.09.2018

Tag der mündlichen Prüfung: 25.10.2018

Darmstadt 2018

D 17

Acknowledgments

I would like to thank first Pr. Beatrix Suess to have given me the opportunity to perform my PhD in her laboratory. These three years were for me a great experience where I could acquire, expend and share my knowledge with a lot of very interesting scientists all around Europe. None of that could have been possible without this opportunity that I was given.

I would also like to thank Pr Alexander Loewer for his kind acceptance for the second opinion on my doctoral thesis.

My acknowledgments also go to Pr. Matthias Heinemann and all the participating professor of the MetaRNA project but also to the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie grant agreement No 642738, MetaRNA and from the European Union's Horizon 2020 to have given me the opportunity to obtain a PhD position inside of this ITN. Thanks a lot to all the ESR fellows of the MetaRNA, meetings with you were always source of inspiration and also pleasant moments, I wish you all success and hope to see you soon.

A huge thanks to the people of the Suess laboratory that made these three years of working a nice journey by having a friendly and enjoyable working days. Special thanks to Thea Lotz for helping me a lot when I just came to Germany. You helped me so much to have a good start in Germany, but also by answering all my stupid questions about German and everyday problems for three years without complaining and for that I would never say enough thank you. I would also thank Dr. Adam Mol for sharing these three years of PhD in the MetaRNA consortium and all travel we made through Europe, you definitively made these trips a nicer experience. A special thanks to also Britta Schreiber, Cristina Bofill-Bosch and Michael "Michi-Mickael" Vockenhuber for your help, your patience with me but also your general good mood that made a nice working atmosphere in the lab. I would also thank Dunja for all the help she provided all along my PhD. Finally, I would like to say a huge thanks to Dr. Florian Groher. Working with you was one of the most enriching (or enrichment?) experience in my working life. I learned a lot with you, all the exchange/meetings/discussion we got during these three years were really fascinating and most of the work I performed would have been much harder to achieve without you. You made me a better scientist, and for that I can only say "Danke schön". I also thank all the other people that belongs to Suess laboratory for making this lab a nice place to work.

I want to say thank you to Daniel Torka and Sandra Ollivaud for the help they provided me during their internship. Sharing with them my knowledge and teach them my work was definitively one of the most enriching experience during my PhD. Part of this work is a credit to you and I wish you good luck for the future.

A very strong and manly thank to Francois-Xavier "gros lard" Lehr. You were a great support during all these three years, as well at work but also outside. I think you are the one that save me from dementia by supporting me all this time, sharing time and effort at the gym, having philosophical discussion on the balcony, for making me meet very interesting people, especially Cassandre that I would also like to thank for her support, but also for making me learned so much and making my French a little bit less bad.

My acknowledgments also go to people that supported me from France. For sharing the PhD experience together even at 1000 km distance away but also for the holidays we spent that helped me to relax and cheered me up and the support they gave me, I would like to thank Geoffrey Prévot and Pierre Vigié and also wish you guys good luck for your coming PhD defence. A thank goes also to my basketball friends, our annual meeting in Hagetmau was for me a nice fresh air breath in this PhD.

Finally, I would also thank my parents, my sister, my grandparents and generally my family for their huge support all along this journey. None of this could have been possible without them.

Table of Contents

1. Zusammenfassung

RNA ist ein Schlüsselmolekül für alle lebenden Organismus. Eine wichtige Aufgabe der RNA in der Zelle ist die Regulation der Genexpression. Eine der Strategien, die von der RNA zur Kontrolle und Anpassung der Genexpression an die Umgebung der Zelle verfolgt wird, ist durch Riboswitche, die die Genexpression als Antwort auf die Gegenwart eines bestimmten Moleküls aktivieren oder hemmen können. Riboswitche bestehen aus zwei Teilen, eine sensorische Domäne, in der die Bindung des Zielmoleküls erfolgt, genannt auch Aptamerdomäne und der Expressionsplattform, die die Genexpression beeinflusst. Basierend auf diesem Prinzip konnten Wissenschaftler künstlich synthetische Riboswitch mit *in vitro* selektierten Aptameren unter Verwendung einer Technik namens SELEX entwickeln. Bislang wurden mehrere synthetische Riboswitches entwickelt, aber es sind immer noch einige wenige im Vergleich zu natürlichen Riboswitches. Eine der Ursachen ist nicht das Fehlen von Aptameren, die ein Zielmolekül erkennen, sondern dass die meisten dieser Aptamere nicht für den Aufbau von Riboswitchen geeignet sind. Tatsächlich ist die Mehrzahl der *in vitro* selektierten Aptamere nicht in der Lage, die Genexpression zu regulieren. Hierzu ist es nötig, dass die Aptamere bei Ligandenbindung eine Konformationsänderung erfahren. In der klassischen SELEX wird jedoch nur auf Bindung und nicht auf Konformationsänderung selektiert. Vor kurzem wurde eine neue SELEX-Methode entwickelt, Capture-SELEX genannt, die die Auswahl von strukturschaltenden Aptameren gegen kleine Moleküle in Lösung ermöglicht, allerdings für ssDNA und nicht für RNA.

Der erste Teil dieser Arbeit beinhaltete die Entwicklung und Anpassung der Capture-SELEX für RNA. Eine vollständige Anpassung des Protokolls war erforderlich, um diese SELEX nun mit RNA durchzuführen. Weiterhin wurden einige Optimierungsschritte durchgeführt, um die Effizienz zu erhöhen, die zu der ersten erfolgreichen Anreicherung von Aptameren führte, die Neomycin erkennen. Verbesserungen wurden hauptsächlich mit dem Ziel vorgenommen, eine bessere Elution spezifischer Binder zu erreichen, aber auch den Anteil unspezifischer Sequenzen während der Selektion zu verringern. Hierzu wurde ein Hitzeelutionsschritt eingeführt, um schwache Binder zu entfernen, die Zeit der spezifischen Elution verkürzt, um nur Aptamere mit einem schnellen k_{on} auszuwählen und das Design der Bibliothek optimiert. Dank dieser Modifikation konnte die Selektion gegen verschiedene Arten von Molekülen erfolgreich durchgeführt werden und ihre Charakterisierung zeigte die Entwicklung von Aptameren mit hoher Affinität und Spezifität.

Da das Hauptziel dieses Projekts die Entwicklung synthetischer Riboswitches war, wurde die mit Paromomycin angereicherte SELEX-Bibliothek zum Screening nach *in vivo* aktiven Sequenzen verwendet. Nach dem Screening von ca. 1200 verschiedenen Kandidaten wurden mehrere Treffer gefunden und ein Kandidat für das weitere Engineering ausgewählt. Von diesem Kandidaten wurde die Sekundärstruktur durch inline Probing analysiert und eine Ligandenbindung durch ITC-Messung von K_D=21 nM bestimmt. Eine hohe Spezifität gegenüber dem Target konnte sowohl in *vitro* als auch *in vivo* gezeigt werden. Mit diesem Kandidaten konnte nicht nur die Expression von GFP, sondern auch von mCherry regulieren werden. Der Kandidat wurde anschließend für die partielle Randomisierung verwendet, um nach verbesserten Mutanten zu suchen. Durch die Kombination mehrerer dieser Mutationen konnte ein Paromomomycin-spezifischer Riboswitch mit einer 8,5-fachen regulatorischen Aktivität entwickelt werden. Basierend auf der Strukturanalyse wurde ein Teil des Riboswitches mit dem Neomycin-Riboswitch für die Erstellung einer Boolschen NOR Gatters kombiniert, was zu einer 30-fachen Regulation führte.

1. Summary

RNA is a key molecule for living organism in all kingdom of life. A major task performed by RNA is the regulation of gene expression. One of the strategies adopted by RNA to control and adapt gene expression to the environment of the cell, is the use of riboswitches that can activate or inhibit gene expression in response to changes in the presence of a specific molecule. Riboswitches are composed of two domains, one where binding of the target happens is called the aptamer and the other domain is called expression platform. Based on this principle, scientists could engineer synthetic riboswitches using *in vitro* selected aptamers by a technique called SELEX. So far, several synthetic riboswitches have been developed but they are still a few compared to natural riboswitches. One of the causes is not the lack of aptamers that recognize a target molecule, but that most of these aptamers are not suitable for riboswitch engineering. In fact, the majority of in vitro selected aptamers are not able to regulate gene expression. It requires the aptamers undergo a conformational change during ligand binding. In classical SELEX, however, only binding and not conformational changes are selected. Recently a new technique was developed for DNA, called Capture-SELEX allowing the selection of structure switching aptamers against small molecules in solution.

The first part of this work included the development and the adaptation of Capture-SELEX for RNA. A complete adaptation of the protocol was needed to perform this SELEX with RNA. Some improvements were designed compared to the original protocol to increase efficiency that lead to a first successful enrichment towards neomycin. Improvements was mostly done aiming a better elution of specific binders but also the decrease of non-specific recovered sequence during the selection. A heat elution step was introduced to remove weak binders, shorten the time of specific elution, select only aptamers with a fast k_{on} and optimize the design of the library. Thanks to these modification, selection against different types of molecules could be successfully achieved and their characterisation revealed the development of aptamers with high binding affinity and specificity.

As the main goal of this project was the development of synthetic riboswitches, the paromomycin enriched SELEX library was used to screen for *in vivo* active sequences. After the screening of about 1200 different candidates, several hits were found and one candidate was selected for further engineering. he secondary structure of this candidate was analysed by inline probing and a ligand binding was determined by ITC measurement of K_D=21 nM. High specificity towards its target could be shown both *in vitro* and *in vivo*. The candidate was then used for partial randomisation to search for improved mutants. By combining several of these mutations, a paromomycin specific riboswitch with 8.5-fold regulation could be engineered. Further engineering was performed based on the

structural analysis to exchange a part of the riboswitch with the neomycin riboswitch for the assembly of an *in vivo* NOR Boolean logic gate resulting in 30-fold regulation.

2. Introduction

2.1. Synthetic biology

According to its name, synthetic biology seems to be only a spin-off of biology. However, the definition of synthetic biology can differ between biologists and engineers. On one hand it can be consider as: “an emerging discipline that uses engineering principles to design and assemble biological components” (1) or on the other hand as: « a new emerging scientific field where ICT, biotechnology and nanotechnology meet and strengthen each other » (2).

Based on the knowledge of molecular biology, synthetic biology unites different disciplines of science going from nanotechnology to immunology and connect existing concepts between engineering and biology. Synthetic biology split itself to classical biology by this engineering aspect to develop in living organism new functions and properties. Emerged in the late 1990’s the development of the field has greatly benefited from the advancement and the discovery of cloning strategies, automated DNA synthesis and sequencing, **H**igh-**T**hroughput **S**equencing (HTS) and the elucidation of the microbial genome. Driven by the purpose of implementing new biological functions to organism or modifying existing systems, the applicability of these artificial construct is rather extensive. Reprogramming stem-cells, development of bio sensors or the production of chemical entities in genetically modified bacteria is just a small list of what is applicable in the field. In the last years, the field went through important breakthrough thanks to the advances in the reprogramming of fundamental cellular processes that operate orthogonally across several species like RNA regulatory tools derived from the Clustered-Regulatory-Interspaced-Short-Palyndromic-Repeats (CRISPR) immune system from bacteria and the expansion of the genetic alphabet by unnatural amino acids provided further opportunities in the framework of cellular engineering (3, 4). Another achievement was performed by the first complete synthesis and assembly of a mini genome in a bacterial cell (5). Yeast was also part of these progress by the first synthesis of a synthetic eukaryotic chromosome (yeast SC 2.0 project) (6). The development and application of synthetic riboswitches in synthetic biology starts to emerge. Recently, riboswitches were used to study the transient dynamics of a “logic gate-like” construct at a single cell level (7) but also to build new devices that can regulate in an ON/OFF way the expression of numerous gene in yeast upon target addition (8).

This doctoral thesis focuses on the development of synthetic riboswitches that can be used for the monitoring of metabolic fluctuations in yeast. Cells are often exposed to fluctuating environments

in terms of, for example, nutrient availability. Such changes require adaptations, which are typically coordinated by complex molecular sensing and regulatory machineries. Biochemical networks are therefore inherently dynamic and adaptive information processing systems. Although many of the players and mechanisms involved are known, we are still far from an actual quantitative understanding of the complex orchestration of these processes. To tackle these huge problems, the development of synthetic biosensors such as riboswitches, in an integrated or transient manner can help for the development new, efficient methods for efficient parameter inference, model selection and uncertainty quantification for complex biochemical network models, as well as develop new ways to model metabolism.

2.2. Nucleic acid aptamers

Aptamers are short synthetic RNA or DNA oligonucleotides sequences which are able to bind a broad range of targets with high affinity and specificity. The type of target is going from small molecules to proteins, viruses or even whole cells. Based on their nucleotide composition, they are able to fold in complex three-dimensional structures allowing them to form binding pockets like their proteins counterparts.

2.2.1. Three dimensional structure of RNA structures

RNA has the unique feature to form high complex three dimensional structures compared to DNA. This characteristic makes them able to bind their target with high specificity and affinity. There are different levels of structures in RNA, the primary, the secondary and the tertiary structure. The primary structure consists in solely of the nucleic acid sequence composition (A, C, G or U). The secondary structure of the RNA is a two dimensional representation of the structure by showing all Watson-Crick and non-canonical base pairing. These interactions are creating more complex structures like stem-loops also called hairpin. Stem-loops consists of two regions of the same strand are complementary in the nucleotide sequence when read in contrary directions. This strand form then a double helix structure that ends with an unpaired loop. Bulges are also another secondary structure created when one side of the stem possess one or more nucleotide than its complementary side. These supplementary nucleotides are not involved in base pairing and are therefore forming single strand structure inside of a hairpin.

Secondary structure is mostly determined by free energy minimization relying on approximations of RNA motif stability depending of the input sequence (9). The most popular tool used is based on the Zuker algorithm (10) that calculate the lowest folding energy of all sequence fragments. Generally,

this prediction gives a good overview of the secondary structure of the RNA. However, these softwares are not able to predict the next level of RNA structure, the tertiary structure.

Tertiary structure involves interactions between secondary structure elements. Loop-loop interaction (also called kissing complexes) happens when single stranded part loop of one hairpin is forming Watson-Crick interaction with another loop of another hairpin. Pseudoknot are also another motif that can be formed by RNA. It consists in nucleotides of a loop that form Watson-Crick interaction with a complementary single stranded sequence. Other complexes such as G-quadruplex (4 guanines interacting with each other through Hoogsteen interaction forming a square planar structure) or three-way junctions (three stem connected together) are just other examples of folding possibilities for tertiary RNA structure (11).

2.2.2. Binding and interaction potential of nucleic acid aptamers

Each nucleobase (A, C, G and T/U), but also the ribose and the phosphate backbone of nucleic acid possess hydrogen binding donor or acceptor atom collections. Hydrogen binding acceptor are either oxygen as a ketone group on a ring or tertiary amine. Hydrogen binding donor are primary or secondary amine groups that contains respectively one or two hydrogen atoms bound to the nitrogen. According to simple valence bond theory, a hydrogen atom should be able of forming only one chemical bond. However, in many cases hydrogen is formally two-valent. This additional bond is called “hydrogen bond”. There are two types of hydrogen bond but here we will focus only of hydrogen bonds with atoms having a higher electronegativity as hydrogen (C, N, O, F, P S, Cl…) which are the most common elements found in nature. These atoms have the capability of forming A-H...B hydrogen bonds (Π-hydrogen bonds) if B has an unshared pair of electrons. Dissociation energies spread in three orders of magnitude (about 0.2-40 kcal/mol). Within this range, the electrostatic, covalent, and dispersion contributions vary in their relative weights and makes the nature of the interaction variating. The hydrogen bond possesses wide transition regions that merge continuously with the covalent bond, the van der Waals interaction, the ionic interaction, and also the cation-Π interaction (12).

RNA or ssDNA are both polar molecules and globally negatively charged due to their phosphate backbone and therefore allow dipole-dipole interaction or ion-dipole interaction. A dipole-dipole interaction is an electrostatic interaction between a negative side of a polar molecule with a positive side of another polar molecule. To allow this interaction, they need to be close together like they are in a case of aptamer-target binding. However, ion-dipole interactions are more commonly found

between aptamer and their targets mostly thanks to magnesium. Magnesium belongs to the alkaline-earth metals and is mostly found in the biological context as its ion form Mg^{2+}. Magnesium is an essential mineral nutrient (i.e., element) for life (13, 14) and is present in every cell type in every organism. For example, ATP (adenosine triphosphate), the main source of energy in cells, must bind to a magnesium ion in order to be biologically active. What is called ATP is often actually Mg-ATP. As such, magnesium plays a role in the stability of all polyphosphate compounds in the cells, including those associated with the synthesis of DNA and RNA. It has been shown in the literature that magnesium is essential for the binding between aptamers and their target by the formation of an ionic bridge between negatively charged residues (15).

Another bonding force involved in DNA and RNA is pi-stacking. Aromatic pi-stacking is of importance in regulating some mechanisms at the gene level. For example, the recognition of nucleic acid by proteins for regulating gene expression is largely controlled by π-stacking of aromatic amino acids and nucleobases (16, 17). Furthermore, neighbouring base-pair stacking controls the gross geometrical feature of a nucleic acid macromolecule (18, 19). Local nucleotide sequences determine the rigidity of a nucleic acid and the capacity of a segment to fold, kink, or supercoil and is, therefore, responsible for creating the local geometry needed for proper enzymatic recognition (20). The probability of the intercalation of a planar molecule between two stacked nucleic acid base pairs is inversely proportional with the strength of their π-stacking (21). DNA polymer analogues where the backbone sugar residues were removed still have the possibility to form the characteristic R-double helix, with stacked base pairs in the usual Watson-Crick base-pairing scheme (22). While hydrogen bonding plays a major role in gluing the two intertwined helices together, preserving the R-helical structure could be protected by π-stacking interaction in despite the absence of sugar residues. Some evidence suggests that the interstrand hydrogen bonding and the nearest neighbour base π-stacking interaction are not independent, as the stacking of nucleobases influences their hydrogen bonding capacity (23).

2.2.3. Advantages of nucleic acid aptamers over antibodies

All these characteristics explain the versatility of aptamers to interact with a broad amount of target in the same way as antibodies (24). However, aptamers have multiple advantages compared to antibodies. Monoclonal antibodies are laborious and expensive to produce involving high cost processes and screening of a large number of colonies. Additionally, a large scale synthesis is needed nowadays for antibodies, but the performance of the same antibody differs in a different batch. Unlike antibodies, aptamers can be generated with high accuracy and reproducibility *de novo* by

chemical synthesis or by *in vitro* transcription at high yields, with a low batch to batch variation and free from biological contamination (25). They can be as well synthesised or modified with various chemical moieties to increase their resistance to nucleases or to fluorophores for the fabrication of bio-sensors (26, 27). Furthermore, it is well known that proteins are easily denatured and lose their tertiary structure at high temperature. On the other hand, oligonucleotides are more thermally stable and maintain their structures over repeated cycles of denaturation/renaturation. Hence, the greatest advantage of oligonucleotide-based aptamers over protein-based antibodies is their stability at elevated temperatures. Aptamers fold back to their original shape and still interact in the way with their target, whereas antibodies are under heat treatment irreversibly denaturized (28). The storage and the stability of aptamers is cheap and easy as they can last for long time stored at –20°C (29). Finally, aptamers can bind molecules going from very small organic molecules or even ions (30), but also to higher size molecules such as peptides, proteins or even whole cells having K_D from the mM range to the low nM range (30–32). All these characteristics make aptamers ideal candidates for a wide range of application.

Aptamers are often used as sensors. In the case of fluorescence detection, the easiest way is to label the aptamer with both fluorophore and quencher. Using a FRET signal between amine-reactive quencher DABCYL moieties and fluorescein, a device detecting cocaine could be built based on the cocaine aptamer (33). Aptamers are able to change in several different conformations. This ability makes them suitable for the building of probes such as aptamer beacons. Upon target binding, hairpin shape or other more complex rearrangement can allow these probes to detect the presence of the target in solution by using fluorophore-quenchers pairs (34–36). Furthermore, quantum dots, gold nanoparticles (NP) and other nano-materials were studied because of their fluorescence quenching properties in combination with aptamers to replace classical quenchers (37–39). The advantage of gold nanoparticles, due to their properties can be applied for the development of colorimetric sensors, removing the necessity of expensive tools to have a readout but only with naked eyes. The aptamer placed in the surrounding of the NP prevent the gold to aggregates thanks to its highly negatively charged backbone, but upon target detection, aptamers are removed from the particles because of their structure switching capacities and cause the gold to aggregate in bigger particles that lead to a change in colour (40). This method could be applied successfully to detect ATP, cocaine, Pb^{2}+, and K^{+} (41).

2.2.4. Application and potential of aptamers

A plethora of other aptasensors have been developed in combination with several analytical equipment. Either with the use of Surface Plasmon Resonance (SPR) or with other microfluidics devices (42, 43) but as they detect changes in mass and optical properties, these methods are not optimal for the detection of small molecules (organic molecules, toxins and metabolites) (44, 45). A variant of the ELISA test has been developed with aptamers. ELISA is one of the most used clinical test to detect with high sensitivity any protein or peptide. An aptamer-linked immobilized sorbent assay (ALISA) was introduced by Kiel's group. They demonstrated the feasibility of this method via a comparative study with ELISA using an antibody (46). Another alternative to ALISA is the use of lateral-flow assay. In this case, the aptamer is coupled to AuNP and forming a coloured complex. Upon sample addition on the pad, the complex will be carried all along the stripe by diffusion. If the sample contains the target, aptamers will bind to it and will not be able to bind on the test zone due to their folding change. The complex will instead migrate to the control zone where all the NP will be accumulated and one band will be visible.

The application of aptamer is not limited to *in vitro* sensing but also can be applied *in vivo*. In 2012, the Jaffrey's lab developed a system that will lead to a broad range of aptamer-based sensors (47). Based on the GFP fluorescence mechanism, a synthetic ligand called DFHBI (or spinach) was synthesised and an aptamer was developed against this molecule. It appeared that upon RNA binding, the fluorescence of the molecule is activated. This result made the development of hybrid aptamers possible, where the folding of the first aptamer (in this case, the sensor part) will allow the folding of the spinach aptamer and make it glow, creating *in vivo* producible sensor ((48–50)).

Application of aptamer is though not only limited to the sensing field. Considering their numerous advantages compared to antibodies, the development of aptamer applied in health became a major axe of research in the aptamer community. The very first aptamer used as a drug, Macugen, is already commercially available for age-related macular degeneration on the form of a single strand nucleic acid pegylated aptamer specific to VEGF165 which has a major role in angiogenesis and permeability (51). Following the example of the Macugen, several aptamer based drug such as AS1411 (a nuclein-specific aptamer) for acute myeloid leukemia (52), ARC1779 (a von Willebrand factor-specific aptamer) for carotid artery disease (53), and NU172 (a thrombin-specific aptamer) for anticoagulation, are currently tested in clinical trials (54). Aptamers as drug delivery system were also developed based on the selection of candidates that could bind to internalized cell surface receptors. One of the example is an aptamer binding the prostate-specific membrane antigen was

coupled to doxorubicin, an anti-cancer drug, onto the aptamer strand. This drug-loaded aptamer complex was able to carry successfully doxorubicin into prostate cancer cells (55).

2.3. SELEX – establishment, optimization and diversification

2.3.1. Development and concept of SELEX

As aptamers are synthetic nucleic acid sequences, they have to be develop in the laboratory. The standard process to develop aptamers is called SELEX. SELEX stands for Systematic Evolution of Ligands by EXponential enrichment. In the SELEX, aptamers are selected from a synthetized library containing a random region between 20 to 60 random nucleotides flanked by two primer binding sites on each side for reamplification after the selection. In this work, the size of the RNA library used is 123 nt using a randomization of 50 nt, achieving a 4^{50} theoretical maximal diversity. This number is equivalent to approximatively $1{,}2*10^{30}$ different sequences or 2 Mmol (10^6) of RNA. With a molecular weight of around 40.000 g/mol the total weight of nucleic acid would be around 80.000 metric tons or $2*10^{10}$ liter of oligonucleotide at 100 µM (standard delivery concentration when order to a furnisher). This equivalent to two times the weight of the "Charles-de-Gaulle" aircraft carrier boat or 8000 olympic swimming pool of nucleic acid which is quantities that are not manageable in the lab. In practice, it has been proved that around 1 nmol of RNA per selection is enough (56, 57). This shows that not all the 50 nucleotides are involved in the binding to the target. The major part of the sequence is responsible for the formation of the secondary structure of the aptamer meaning that these nucleotides can be variable as long as the folding does not change and the base-pair are kept. The advantage of long motif such as 50 nt is that a small binding motif can be presented in different folding possibilities, such as stem loop, bulge or three-way junction.

The original protocol was developed in the early 90's by three different working groups in parallel (58–60). The selection process is split in several step. The first one is the synthesis of the library. For the first round the library is synthesised chemically in an amount big enough to achieve a theoretical diversity of at least 10^{15} different sequences. The library is afterwards amplified to generate enough copies of each sequences and afterwards converted in single-strand oligonucleotides (either by separating the two strands for DNA or by *in vitro* transcription for RNA). Once the library is generated, the DNA/RNA are folded by denaturation/renaturation and incubated with the target molecule which is immobilized on a matrix. During this step, aptamers able to recognize the target will be able to bind and form stable interaction. Using different type of methods, depending of the nature of the matrix used, the unbound nucleic acids are washed away by using selection buffer and removed from the pool. Aptamers with a good enough affinity to the target will go through the

washing steps and remain on the target. These aptamers are afterwards recovered by a non-specific elution from the matrix (heating, chelation of the ions necessary for the interaction or harvesting of the matrix) or a specific elution (addition of an excess of target in solution to force aptamers to undock from the originally bound target). Once recovered, these aptamers are amplified and used again in another selection cycle.

2.3.2. Optimization and diversification of SELEX strategies

Through the years, SELEX never stop to evolve in order to create a more complete and reliable method. Some basics ameliorations were made at the beginning such as the introduction of the Counter-SELEX (61, 62). This technique was established to segregate in the library non-specific binders to the target that could also bind an unwanted closely related target. The counter-elution step happens as follow: The counter-target is implemented into the matrix and will fish out the non-specific binders. Furthermore, ameliorations on the libraries were performed, such as the using modified bases (26) and nuclease resistant backbone (63–65). These ameliorations were making to aptamers more resistant to degradation on one hand, and on the other hand allowed aptamers to have more interaction possibilities thanks to the introduction of chemical moieties not originally present on the nucleic acids. The drawback of these modification is that no organism is able to produce these sequences *in vivo* as they are fully artificial. Also, special enzymes are necessary to treat with these modifications during the SELEX (66). Finally, SELEX method expended to a wide range of support and the method got diversified. Cell-SELEX is one of the most known example. This technique is not using anymore a matrix to bind the target protein but directly living cells that are expressing the protein of interest directly on the surface of the cell (67). Thanks to this technique, aptamers could be selected for membrane proteins (68) which are very difficult to produce *in vitro* and if succeeded, are probably not in the folding they should be at the surface of a cell (69). Capillary Electrophoresis (CE) SELEX is another alternative to perform SELEX. This technique has the advantage to realize the selection without any immobilisation of the target. Indeed, based on the mobility shift observed when there is a formation of a complex aptamer-target, the binding sequences are migrating at a different rate allowing the separation and collection of high-affinity aptamers from the non-binding sequences (70). This process was the first step towards a faster SELEX protocol. Actually, all the protocols mentioned beforehand have the drawbacks of being longer and time consuming. However, CE-SELEX requires a specific equipment and an adapted formation to be performed.

2.3.3. Capture-SELEX – an optimized method for free target selection

In 2012 a paper was published showing a new way to perform SELEX on DNA called Capture-SELEX (71). This strategy was developed to allow the selection of aptamers against target that was not suitable for classical immobilisation like small molecules. Indeed, the immobilisation of the target requires the formation of a covalent bond between the target and the matrix. In the most often cases, either a primary amine or a carboxyl group is used to create this bond through a chemical reaction. This way of action is suitable for most of the proteins because they have enough amino acids that contains these reactive groups on the side chain of some amino acids. Furthermore, thanks to its size, most of the protein surface will still remain available to interact with potential aptamers during the SELEX. However, this is not the case for small molecules (antibiotics, metabolites…). Indeed, small molecules have a limited amount of reactive group on their surface (sometimes none), requiring often the usage of hard chemistry to perform the immobilisation (8, 72). Leading often to either a poor yield of immobilisation or even in worst case the complete degradation of the target that couldn't undergo the immobilisation conditions. In the Capture-SELEX strategy, the roles are inverted. The DNA is immobilized and the target is free in solution. For this purpose, a special library had to be developed. The DNA pool in addition of the primer binding sites and the randomised region, contains a small conserved sequence, placed inside of the randomised part. This sequence is called the docking sequence (DS). This DS will be used has an anchor for another small oligonucleotide which is complementary to the DS, called the capture-oligonucleotide (CO). The CO is derivatized on its 3' side with a linker terminated by a biotin. Thanks to this biotin, the complex CO-DNA library can be immobilized on magnetic streptavidin-coated beads, forming a complex were the library is immobilized on a magnetic support. As all DNA are not able to interact with the CO due to different folding possibilities, some wash steps are necessary to remove unbound DNA from the complex beads-pool. After removing these sequences, the elution of the target binders has to be performed. For that the target mixed with selection buffer is added directly on the beads. Next, only aptamers that undergo a change in their structure upon target binding will be set free from the magnetic beads and can be recovered by applying a magnet on the tube to separate them and reamplify them for the next SELEX cycle.

This technique contains several advantages. The most important one is that Capture-SELEX is suitable for all kind of small molecules as long as they can be soluble in selection buffer at a high enough concentration. Indeed, the target only need to be mixed with the beads to allow the elution of aptamers. This technique is as well faster and straightforward; all the steps can be tuned very

easily. Another big advantage is the selection of only structure switching aptamers. One can see this feature as a drawback. Indeed, aptamers that do not change structure upon ligand binding will stay on the beads and therefore not be eluted. Hence, some interesting candidates can be lost during the selection. However, this structure-switching characteristic is very interesting for downstream applications. When we take a look at biosensors there are different possibilities. One can be the coupling of the aptamer with another aptamer for a fluorescent dye such as spinach aptamer or mango aptamer. Indeed, these aptamer have the particularly to exhibit fluorescence only upon target binding. When couple to a structure switching aptamer for a metabolite for example, the folding of the unbound state of this aptamer will not allow the formation of the dye aptamer and therefore no fluorescence will be observed. Nevertheless, if the metabolite is detected in the medium, the metabolite aptamer will change structure and consequently will correctly fold the dye aptamer that will be able to bind its target and to fluoresce. The other possibility is the development of a synthetic riboswitch that can be exploited as internal bio sensors in a cell.

2.4. Riboswitches, natural and synthetic regulators with diverse applications

RNA was always thought as only a messenger of genetic material between DNA and protein. Although proteins fulfil most of the requirements that biology has for structural and functional components such as enzymes and receptors, RNA can also serve in these capacities. For example, RNA has sufficient structural plasticity to form ribozyme (73, 74) and receptors (75, 76) elements that exhibit considerable enzymatic power and binding specificity. In the last decade, the discovery of many different types of regulatory RNA has been made such as miRNA, other non-coding small RNA and riboswitches.

The term riboswitch regroups all RNA-regulating elements that are sensitive to the input of a target molecule, most often a metabolite, and transmit this input into a structural rearrangement in the architecture of the mRNA, regulating the expression of the downstream gene. Riboswitches have been discovered in all kingdoms of life. Originally, researchers tried to elucidate the regulatory mechanism of the B1, B2 and B12 biosynthesis (77). After having search in vain a protein based regulatory feedback control, the involvement of mRNA sequences acting as regulator was speculated. It has been found that the previous quoted vitamins were interacting with their respective mRNA and consequently confirmed the regulatory role of RNA in metabolite biosynthesis (78, 79). Riboswitch consist solely of RNA that sense their target in a preformed binding pocket and undergo a structure-switching upon target binding, whereby one of the two conformations

interferes with gene expression. Hence, riboswitch perform both sensor and effector functions, showing that a protein cofactor is not always required for regulation.

2.4.1. Natural riboswitches

Riboswitches is a class of genetic regulatory elements that are widely distributed throughout the bacterial world. RNA located in the 5' UTR of a mRNA is used as a molecular switch to control transcription termination, translation initiation or mRNA processing through changes in their three-dimensional structure depending on the interaction with a specific cell metabolite such as vitamins, amino acids and nucleotide analogues (80, 81). Here, they participate the maintain of the steady-state between educts and products consumed and synthesized for the biogenesis of a metabolite, controlling its flux and abundance with a low metabolic cost. While the binding domain (or aptamer part) is highly conserved constant to keep its binding properties, the expression platform has been declined through evolution in several alternatives. Typically, transcription controlling riboswitches are either Rho-independent terminator (formation of a stable stem loop that makes the RNA polymerase II stall and fall) or anti-terminator hairpins (sequester the terminator sequence and allow the elongation of the mRNA). Translation regulating riboswitches involve mostly the Shine-Dalgarno sequence by either sequester it upon ligand binding (translation initiation inhibition) or by setting it free upon ligand binding (translation initiation activation) (80).

2.4.2. Synthetic riboswitches

The design and implementation of riboswitches give the opportunity of handle and control the expression level of any gene *in cis* with a low metabolic cost. Their nucleotides composition and their property to respond in an external *stimuli* make their engineering suitable for plug-and-play approaches. Based on this principles, synthetic riboswitches have been engineered combining RNA aptamers domains with regulating domains.

In prokaryotes, the ribosome binding site (RBS) recruits the ribosome and place it in front of the start codon. Translation regulation by synthetic riboswitches was mostly done by their insertion at proximity of the RBS to change the ribosome accessibility to this sequence. Upon ligand binding, the structure-switching properties of the synthetic riboswitch will sequester the RBS and therefore block the downstream translation. The theophylline aptamer (82) was used in this purpose by being inserted nearby the RBS and flanked with randomized nucleotides. This library was cloned in front of the chloramphenicol resistance gene and candidates were selected candidates that can activate the gene when the target is present, but where no gene is expressed when the wrong target or no

target is added. Thanks to this approach, a 11-fold increase ON-switch was engineered. This riboswitch was furthermore used in front of numerous gene (83, 84) and in several hosts (*Streptomyces coelicolor* (85), *Francisella novicida* (86)…) in follow-up studies and could in the best case reach a regulation factor up to 200-fold in the case of a gene regulation in *Synechococcus elongatus* (87). The theophylline aptamer was also used for the design of riboswitch harnessing transcription termination. A complementary spacer sequence of the 3' end of the aptamer was added between the aptamer domain and the poly-U stretch. Without theophylline, the spacer generates a stem loop with aptamer domain and terminates the transcription. Upon theophylline addition, the spacer the aptamer folds properly and the spacer region is set free, allowing the transcription to finish (88).

In eukaryotes, other strategies have to be develop in adaptation to a whole different machinery for gene expression. The first aptamer that could regulate gene expression in eukaryotes was made of two copies of the Hoechst dye aptamer that were placed into the 5' UTR of a luciferase gene showing a 10-fold down regulation upon target addition by controlling the cap-dependent translation initiation (89). Since the two last decades, several synthetic riboswitches were developed from SELEX enriched libraries showing similar regulations. First, a tetracycline dependent riboswitch was developed. When inserted into the 5' UTR of several genes, upon tetracycline addition, the expression level was decreased 6-fold. Depending of the location of the insertion, the riboswitch was regulating either the access to the cap-structure by the small ribosomal subunit when placed directly downstream of the cap structure or blocking the scanning of the ribosome when placed close to the start codon (90). Based on the success of this experiment, Weigand *et al* developed a new synthetic riboswitch towards neomycin B from an enriched *in vitro* selection (72) by a combined approach of *in vitro* selection and *in vivo* screening (91). Recently, Groher *et al.* (8) applied this approach and established a pipeline with success for the development of synthetic riboswitch from *in vitro* selection to optimized *in vivo* screening protocol to find and optimize *in vivo* active riboswitches. Part of this pipeline will be use in this thesis. The ciprofloxacin-riboswitch was demonstrating to be a typical plug-and-play device for any gene regulation in *Saccharomyces cerevisiae* by its ability to regulate the expression of several different genes coding for GFP, kanamycin resistance or URA3 synthesis by regulation of the ribosome scanning.

Splicing regulation is an interesting alternative possible only in eukaryotes. During this step, pre mRNA is modified by the removal of introns sequences by the spliceosome to be translated in active proteins. Preventing this step to happen will lead to the production of inactive proteins. When

placed in proximity of crucial splicing elements such as the 5' splice site, the branch point or the 3' splice site, riboswitches have the possibility to control their accessibility to the spliceosome. A successful riboswitch responding to tetracycline could be engineered to control access to 5' splice site, reaching a regulation factor in yeast of 16-fold (92, 93). The tetracycline riboswitch assumes a three-way junction structure where the target binds in the centre of this structure. The riboswitch is closed by a 4 nt stem. In this experiment, the closing stem of the switch was elongated on its 5' side with the 5' splice site. Several construct were designed where the 5' splice site gets either before the closing stem, partially integrated in the stem or completely part of the stem. The construct showing the higher regulation factor was completely involving the 5' splice site in the closing stem, allowing the access of the U1 snRNP only in the absence of tetracycline but sequestering the splicing site upon target binding leading to unspliced mRNA and consequently to non-active GFP. The same approach was tried by adding the branch point region (also necessary for the binding of the Splicing factor 1) to the 3' side of the closing stem. However, branch point regulation was not active in yeast but was successfully engineered in HeLa cells (94).

2.5. Target molecule used for SELEX

The need for *in vivo* active RNA biosensors in the field of synthetic biology, strain producer optimization or for better understanding of cell metabolism has become bigger over the years. Detection for better antibiotic producer in antibiotic producers is a long and expensive process in terms of sample preparation and analyse of the result. The same challenges appear about monitoring the metabolism on a single cell level. An alternative option would be the use of RNA biosensors as riboswitch. If couple to a fluorescent reporter gene, the concentration of a given metabolite can be followed just by a simple measurement. Therefore, SELEX against these classes of molecules was tried.

2.5.1. Antibiotics

An antibiotic is an active molecule possessing antimicrobial activity and used in the treatment or the prevention of bacterial infections. Antibiotics act on the infection by either inhibiting the bacterial growth or killing bacteria cells. Antibiotics can be produced naturally by fermentation in living organisms (penicillin), can be semi synthetic based on a natural antibiotic but derivatized with a moiety to increase their spectrum of use (ampicillin) or completely synthetic such as the chloroquine. For industrial microbiology fermentation is processed in huge containers containing hundreds of thousands of litres of growth medium. Here, all parameters are important (oxygen concentration, temperature and nutrient level for example) and needs to be controlled. As

antibiotics are secondary metabolites optimization of productions depends of numerous parameters that can make laborious to find the optimal strain for optimal production. Detection for better antibiotic producer in *fungi* (in which e.g. penicillins are produced) is a long and expensive process in terms of sample preparation and analyse of the result. The same challenges appear about monitoring the metabolism in a single cell level. An alternative option would be the use of RNA biosensors as riboswitch. If coupled to a fluorescent reporter gene, the concentration of a given metabolite can be followed just by a simple measurement.

2.5.1.1. Aminoglycosides

Aminoglycosides are highly potent, broad-spectrum antibiotics with many desirable properties for the treatment of life-threatening infections (95). They were first discovered in 1944 by the discovery of streptomycin produced by the actinobacteria *Streptomyces griseus*. Thirty years later, semi-synthetic aminoglycosides such as dibekacin or amikacin were demonstrated active against strains that developed a resistance to the original aminoglycosides (96). Their mode of action consists in impairing bacterial protein synthesis through binding to prokaryotic ribosomes. Aminoglycosides perturb the elongation of the nascent chain by damaging the proofreading process controlling translational accuracy, creating misreading and/or premature termination (97). They are mostly uptake in the bacteria by self-promoted uptake process (98). In addition, aberrant proteins produced due to the effect of the antibiotic may be inserted into the membrane leading to higher permeability and therefore further import of aminoglycoside (99). Paromomycin, neomycin and kanamycin A are all natural aminoglycosides discovered in the 1950's in three different Streptomyces strain (*Krestomuceticus* (100)*, fradiae* (101) and *kanamyceticus* (102) respectively). The three of them are highly soluble in water (>100mM), have rigid and flat structure and possess a lot of amino groups and hydrogen bond donor/receptor. All these characteristics make them excellent candidate to interact with RNA and explain their efficiency to target the ribosome.

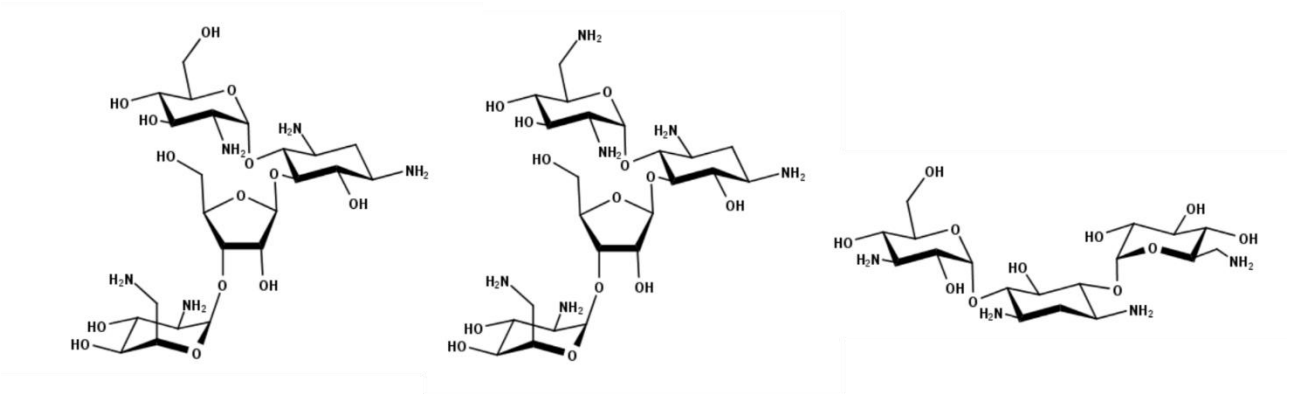

Figure 1.1. Structure of the aminoglycosids antibiotics paromomycin (left), neomycin (middle) and kanamycin A (right).

2.5.1.2. Penicillin

The penicillin derivative antibiotics belong to the family of beta-lactam antibiotic that include penicllin G, penicillin V as naturally produced antibiotic, but also semi synthetic aminopenicillin such as ampicillin and amoxicillin. The penicillin antibiotics are industrially produced using *Penicillium fungi.* Their mechanism of action is inhibiting the biosynthesis of the cell wall of bacteria by blocking the formation of peptidoglycan cross-links in the bacterial cell wall acting as analogues of D-analyl-D-analine. This is achieved through the binding to the DD-transpeptidase that cannot anymore catalyse formation of these cross-links and lead to cell death by causing an imbalance in cell wall production (103). Penicillin antibiotics are industrially produced by fermentation in optimized strain. As optimization is a never ending process, new optimized strains are still under research to produce more antibiotics at lower cost. This is one of the potential application of sensing riboswitch, that could give an easy-to-read output for antibiotic concentrations.

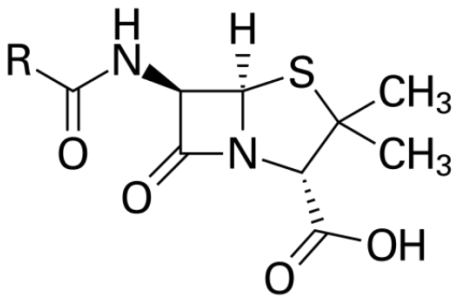

Figure 1.2. Structure of the penicllin antibiotics core.

2.5.1.3. Viomycin

Viomycin belongs to the family of non-ribosomal peptide antibiotics and is one of the essential component in the drug cocktail used to fight tuberculosis. Viomycin like aminoglycosides is produced in an actinomycete, *Streptomyces puniceus*. There are several mode of action for viomycin. First, tuberactimomycins (family of antibiotic in which belong viomycin) are known as inhibitors of the protein synthesis by acting on both ribosomal subunit inhibiting both initiation and translocation (104, 105). Viomycin contains also a guanidine groups that inhibit self-splicing of group one introns by acting as competitor. When Viomycin is bound, the guanosine cofactor that perform the nucleophilic attack the 5' splice site cannot access anymore to the splicing region (106). Due to its high amount of amino groups and hydrogen bond donor and receptor, viomycin can bind to RNA and explaining its efficiency to inhibit prokaryotic protein synthesis and certain form of splicing.

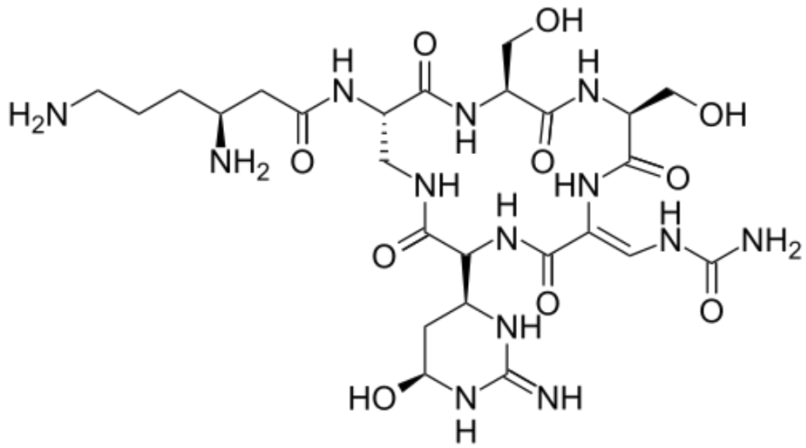

Figure 1.3. Structure of viomycin.

2.5.1.4. Chloramphenicol

Chloramphenicol (CMP) is a bacteriostatic antibiotic from the amphenicols family originally isolated from *Streptomyces venezualae*. CMP is inhibiting the protein synthesis by preventing the protein chain elongation inhibiting the peptidyl transferase activity of the bacterial ribosome. It interacts with two specific residues (A2451 and A2452) in the 23S rRNA of the 50S ribosomal subunit (107). Besides, CMP possess two Cl atoms that are highly electronegative. This particularity increases the probability of attracting other electrons for an interaction.

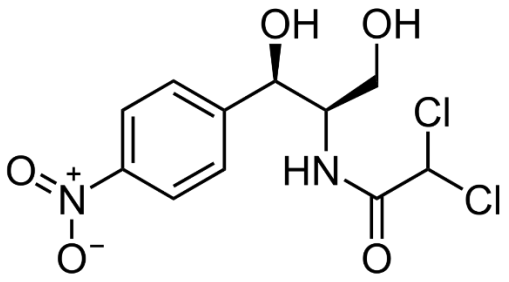

Figure 1.4. Structure of chloramphenicol.

2.5.2. Metabolites

2.5.2.1. Citrate

Citrate is an important metabolite that is found in a critical step at an intersection in metabolism of mammalian cells. It is produced in the mitochondria and used in the Krebs cycle. Citrate and its derivative (oxaloacetate and acetyl-CoA) are used in both normal and pathological process. More than its classical role as metabolic regulator, citrate has been shown in the literature as involved in different processes such as inflammation (108), cancer (109) and insulin secretion (110).

2.5.2.2. ATP

ATP is a ribonucleotide that furnish the necessary energy to chemical reaction of the metabolism from locomotion, to cell division or also to active transportation of chemical species through biological membranes. ATP is also one of the building block for RNA and therefore can interact with nucleic acid through its Watson-Crick interaction side, through the Hoogsteen site, or by pi-stacking. ATP is also made of a ribose that can perform hydrogen interaction and of three phosphates groups in a row that offer also possibilities of interaction especially in complex with magnesium. Monitoring the level of ATP in real-time using an RNA biosensor is the best option to observe the fluctuation of the metabolism.

2.5.2.3. Phospho-enol-pyruvate

Phosphoenolpyruvate (2-phosphoenolpyruvate, PEP) as the ester derived from the enol of pyruvate and phosphate is present in the cell under is anion form. PEP is an important intermediate in biochemistry as it forms the highest-energy phosphate bond found (−61.9 kJ/mol) in organisms PEP is as well involved in glycolysis and gluconeogenesis. PEP is also involved in the biosynthesis of various aromatic compounds in plants but as well in carbon fixation. Furthermore, in bacteria, PEP can be also used as the source of energy for the phosphotransferase system (111).

2.5.2.4. D-ribose-5-phosphate

Ribose-5-phosphate is a metabolite belonging to the pentose phosphate pathway resulting from the isomerization of ribulose-5-phosphate, produced at the last stage of this metabolic pathway. Ribulose-5-phosphate can undergo various isomerizations, transaldolizations and transcetolizations to produce other pentose phosphates as well as fructose-6-phosphate and glyceraldehyde-3-phosphate, two intermediates of glycolysis. Ribose-phosphate diphosphokinase converts ribose-5-phosphate to phosphoribosylpyrophosphate (PRPP), a precursor activated for nucleotide synthesis, such as AMP.

2.5.2.5. Fructose bisphosphate

Fructose-1,6-bisphopshate (FBP) is a metabolite located at the crossing of several metabolic pathways like the glycolysis and the gluconeogenesis. Most of the glucose and fructose metabolised by the cell is at one point transformed in FBP. FBP plays a central role in metabolism regulation. For example, FBP can activate several enzymes from the glycogen metabolism (ADP-glucose phosporylase), from the glycolysis and gluconeogenesis (phosphofructokinase) and from the Calvin cycle and pentose phosphate pathway (RuBisCo oxygenase) but also downregulating some enzymes in the same pathways like glycogen phosphorylase, fructose-1,6-phosphatase and 6PG dehydrogenase (112).

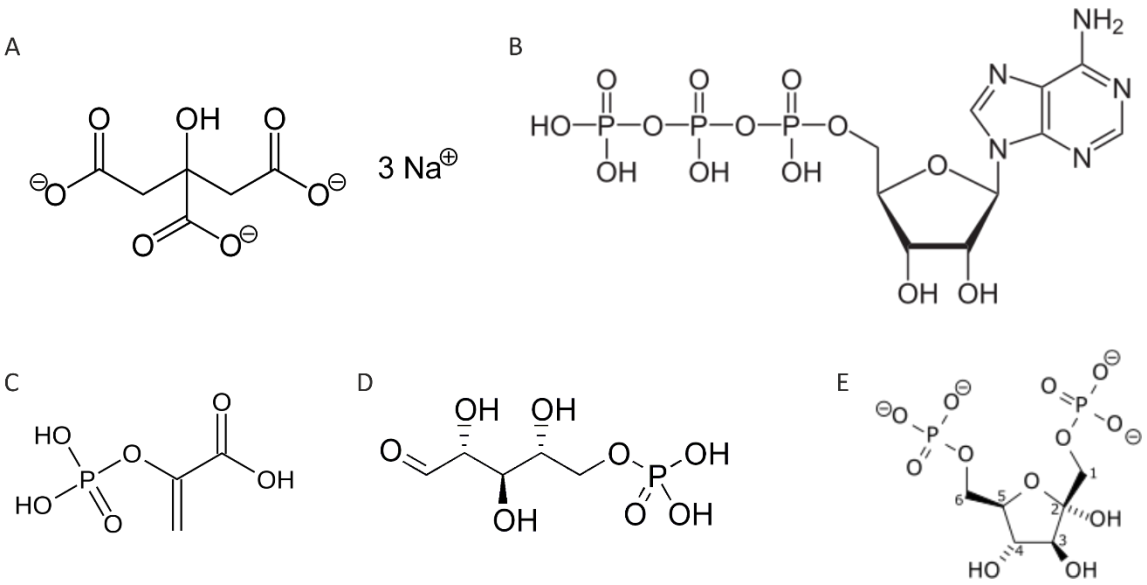

Figure 1.5. Structure of the metabolites citrate (A), ATP(B), PEP (C), D-ribose-5-phosphate (D) and FBP (E).

2.5.3. Glyphosate

Glyphosate (*N*-(phosphonométhyl)glycine) is a non-specific herbicide absorbed by the leaves of a plant. Glyphosate act as an inhibitor of the 5-enolpyruvylshikimate-3-phosphate synthase, an enzyme of the aromatic amino acid biosynthesis. Nowadays, glyphosate is the most commonly used herbicide worldwide due to its high efficiency removing plants. However, due to its massive abuse in agro alimentary industry and the increasing of glyphosate-resistant plants, glyphosate diffused a lot in rivers and lake contaminating a lot of drinkable water. In 2014, an analysis published a possible increased risk of non-Hodgkin lymphoma in workers exposed to glyphosate formulation (113).

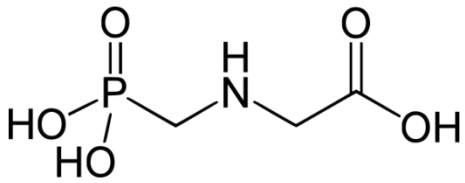

Figure 1.6. Structure of glyphosate

2.6. Aim of the study

Aptamers are the sensing domains for synthetic riboswitches. Development of synthetic riboswitches used as a biosensor offers a broad range of application for example for monitoring metabolic fluctuation on single cell level or for strain optimisation. However, the number of the available tools is very low compared to the amount of published aptamers and only a small subset of aptamers can be used and is used for RNA engineering (24). The supposed theory is that classical SELEX is optimized to select for aptamers with high affinity and specificity, but not for structural rearrangement often supposed to be the key for natural riboswitches. Indeed, all natural and synthetic riboswitches demonstrate a change in their structure upon ligand binding to have an impact on the transcription, translation, splicing or other cellular machinery (114). Therefore, a new SELEX strategy had to be developed, that implements the need for structural refolding.

One aim of this study was to develop, establish and optimize a RNA based Capture-SELEX approach. Capture-SELEX strategy relies on the immobilization of the library by a Watson-Crick interaction between a small capture-oligonucleotide that retain the pool with its interaction with a small complementary sequence on the library (the docking sequence). This feature makes it able to select and enrich only with structure-switching aptamers. Up to know, the protocol is established for DNA Selex only and therefore a complete adaptation for RNA is required because of their inherent differences and the needed changes in the whole process e.g. transcription…. Establishing a protocol

for SELEX includes the proper tuning of many different process parameters. Indeed, the selection protocol is divided in several subparts that all have limitations and drawbacks that need to be overcome and optimized to tend towards a universal RNA Capture-SELEX protocol.

To prove the unique features of Capture-SELEX to find *in vivo* active riboswitches - compared to the classical SELEX approach - a proof of concept *in vivo* screening was performed in baker's yeast to look for sequences able to modulate gene expression upon ligand binding. The optimized RNA Capture-SELEX should then be able to deliver an *in vivo* active aptamer (riboswitch) having a high affinity and specificity towards its target ligand. The discovered riboswitch(es) has to be then optimized using low rate mutagenesis in order to find gain-of-function mutations to end up with a synthetic riboswitch having the same regulation range as all the one already existing.

3. Results

3.1. Development of the RNA Capture-SELEX

The classical SELEX approach is not suitable to any class of ligands. The technique requires the immobilization of the target on a solid support to allow the partitioning between the aptamers (that will bind to the target linked to the matrix) and the non-binding sequences that will flow through and get discard. A lot of small molecules organic molecules such as antibiotics or metabolites but also bigger like protein are not adapted for this treatment. The need for a chemical immobilization by the formation of a covalent bond is often made at non-physiological condition that degrades irreversibly the target. In addition, small molecules which offer less interaction possibilities as protein for example, as they will have one reactive moiety block and unavailable for interaction with the aptamers. To circumvent its drawbacks, an alternative strategy called Capture-SELEX was developed for DNA (71). In this paper, the authors describe a method where the need of chemical immobilization of the target is replaced by the capture of the RNA pool by natural Watson-Crick interaction using a small complementary oligonucleotide as an anchor immobilized on magnetic beads-based support. In short, the DNA library (figure 3.1.) is mixed with a complex of streptavidin derivatized magnetic beads coupled with biotinylated capture oligonucleotide (CO). This CO has a defined sequence which can bind to the complementary docking sequence (DS), a fixed short region located in between the randomized region of the DNA library. The library is flanked by two primer binding sites (PBS) for further PCR amplification.

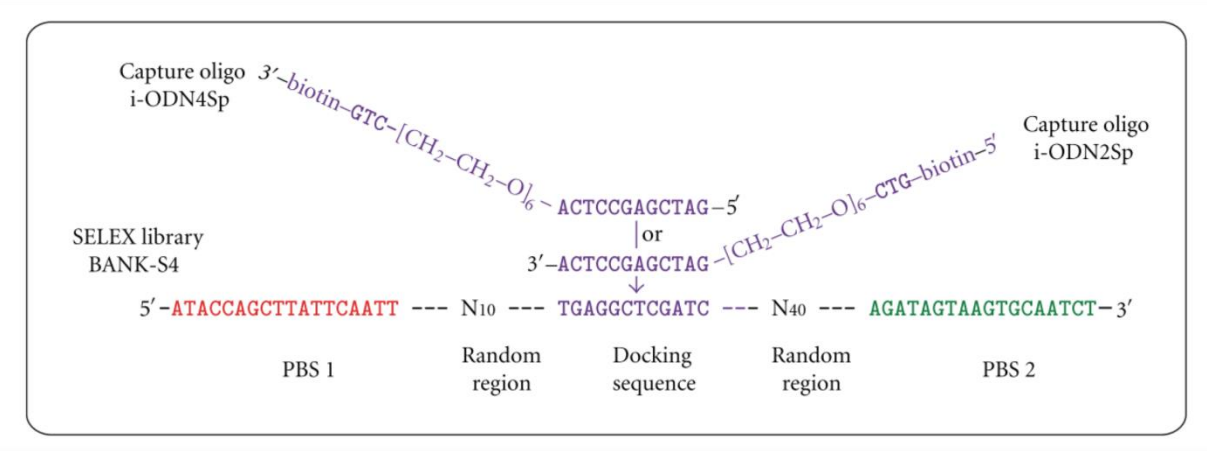

Figure 3.1 Design of the DNA Capture-SELEX library from Stoltenburg et al adapted from (71). The 5' and 3' primer binding sites are respectively in red and green. The DS is in purple and located between a 10 nt long random region and a 40 nt long random region. The complementary CO is attached either at the 5' or 3' side to a hexaethylene glycol spacer followed by three nt and a biotin at its end.

The selection concept is based on the release of specific DNA binding sequences from the CO which shows an affinity to the target and therefore undergo a specific conformational change for binding to the target in solution (scheme of the protocol in figure 3.2.).

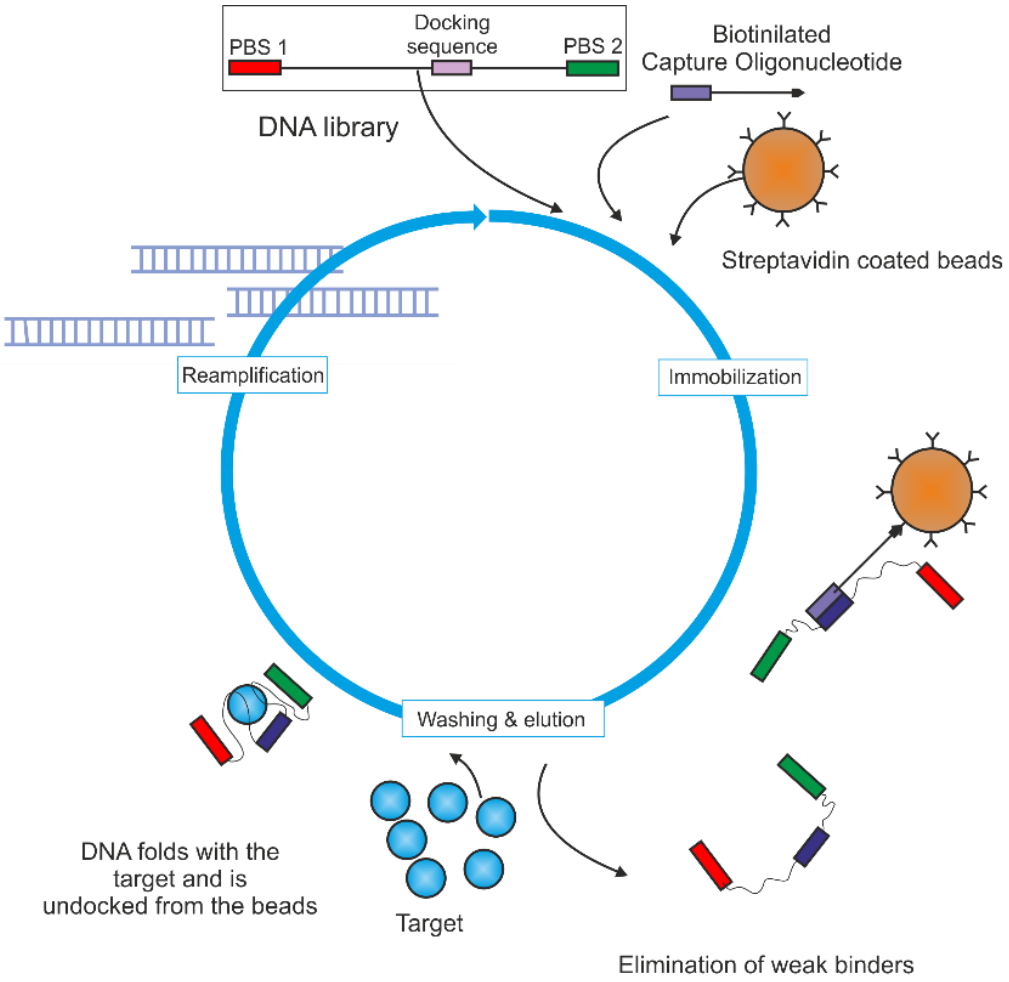

Figure 3.2. Capture-SELEX strategy (71). Schematic representing the DNA Capture-SELEX protocol. In the first step, the library is mixed with streptavidin-coated mixed and a biotinylated capture-oligonucleotide. The library gets immobilized on the capture-oligonucleotide, that is bound to the streptavidin-coated beads. After several washing steps to remove weak binders to the capture-oligonucleotide, the target is added to the beads and sequences that undergo a structure switching upon ligand binding get undock. These sequences are afterwards recovered and amplified for another cycle.

One possible risk of this concept is that not all aptamers which can bind to the target also undergo a conformational change including the DS. These aptamers will not be released and therefore not further enriched. On the other hand, there are more advantages offered by the Capture-SELEX strategy that the classical approach can't provide. The major one is the complete absence of need for chemical immobilization of the target. This hard chemistry often leads to the partial or total degradation of the target by extreme pH or buffer conditions and is blocking one chemical moiety for binding. Another advantage is the speed and the accuracy of the target condition that can be

applied. The target can directly be prepared a few minutes before the experiment to avoid any degradation due to low stability of the compound in the SELEX buffer. The major advantage, however, of this strategy is that the aptamer selected will be only sequences that undergo a structural switch upon ligand binding. This characteristic is one of the main conditions to find a riboswitch. Consequently, if this technique was applied with success on RNA, the probability of finding an *in vivo* active aptamer may be higher than with the classical SELEX approach. In this first part, I will develop how to adapt the original Capture-SELEX protocol to make it suitable with RNA aptamer selection.

3.1.1. Establishment of the protocol

As the original Capture-SELEX was realized with DNA, the whole protocol had to be readapted for RNA. Indeed, RNA can't be heated up at high temperature in the presence of magnesium and the reamplification step must be changed completely. The Capture-SELEX protocol is divided in several steps: First, the library immobilized on the streptavidin-coated magnetic beads by the through its interaction with the CO forming a complex Beads-RNA in solution. Then, the beads are washed, pelleted and the wash fraction containing non-immobilized RNA is removed. This step is repeated several times to eliminate a maximum of unbound RNA. Afterwards, the target is incubated in the presence of the beads. Aptamer that undergo a structure switching upon ligand binding are undocked from the CO. Others sequences are still bound to the beads. Eluted aptamer are recovered and amplified using RT-PCR followed by in vitro transcription. After a purification step, the RNA is ready for the following round.

3.1.1.1. Assembling of the complex

Design of the library. The key factor of the Capture-SELEX is to choose the right docking sequence with the right characteristics (melting temperature, sequence content, secondary structure…). The initial idea was to design a "ready-to-use" pool for splicing control by intron retention based on a riboswitch regulating gene expression by control of the splicing (92). In this paper, the authors developed a riboswitch able to control splicing in yeast by involving splicing regions such as the 5' splice site and the branch point in the closing stem of the tetracycline riboswitch. Thus, once involved in the folding of the aptamer upon ligand binding, the splicing regions will be blocked for the access of the spliceosome (U1 for the 5' splice site and splicing-factor 1 for the branch point). With this strategy, they could reach a 16-fold regulation factor by controlling the access to the 5' splice site. The regulation by controlling the access to the branch point region was, however, only successful in human cells with this riboswitch. As observed in the analysis of the folding prediction of the aptamer

found with DNA Capture-SELEX (figure 3.3. (115)), refolding of the aptamer upon ligand binding involves the DS to interact internally with the aptamer to be able to get release from the CO and be recover.

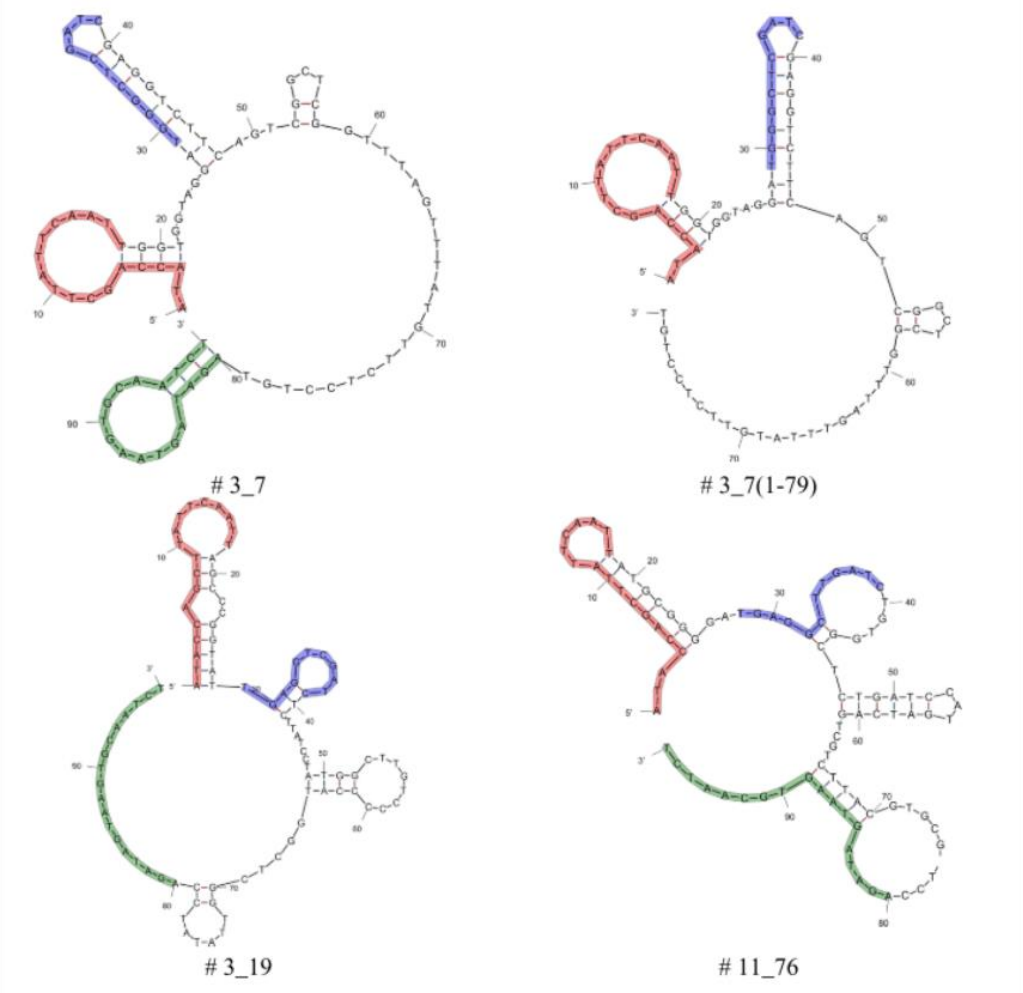

Figure 3.3. Secondary structure folding predictions of DNA aptamers against kanamycin found by capture-SELEX, from (114). The folding of these sequences was performed by means of the free-energy minimization algorith using the web based tool mfold with salt correction. The primer binding site are highlighted in red and green. The docking sequence is highlighted in blue.

Considering this particularity, several designs of libraries were made where sequences important for splicing region were included in the pool as part of the DS. Primer binding sites were designed based on a previously published library (72) and modified to integrate splice sites when needed (figure 3.4.). Based on several introns known from the literature, the intron YPR187W was chosen. The criteria of selection were: no ATG in the sequence, constitutive 5' splice site, an intron length the smallest possible and a strong 3' SS (no AAG).

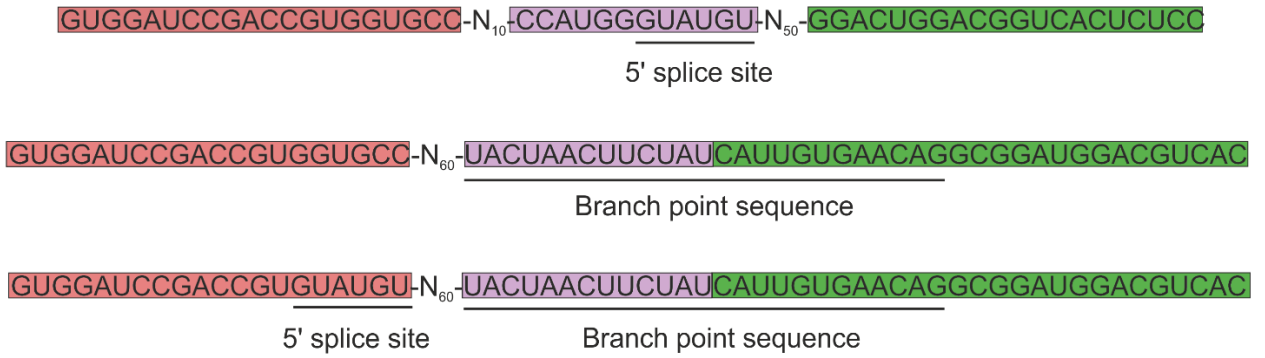

Figure 3.4. Design of the libraries with splice regions for Capture-SELEX. The primer binding site are in red and green boxes. The docking sequence is in purple box. Splicing regions (5' splice site and branch point) are underlined. Random regions are represented by the N and its length by the number written in subscript.

As the docking sequence is involved in the structural switching upon ligand binding, the DS was designed either as part of the 5' splice site (first construct) or 3' branchpoint region (second and third construct). The first one contains the start codon followed by the 5' constitutive splice site in the docking sequence and is flanked by N10 on the 5' side and N60 on the 3' site. The second one contains the beginning of the 3' branchpoint region on the docking sequence with only N60 on the 5' side of the DS. The last one is identical as the previous one with the addition of the 5'constitutive splice site at the end of the 5' primer binding site. Among these three constructs, the one that was estimated the best was the third one. Indeed, by containing both 5' and 3' splice site, it will make a build-in system for cloning and screening for splicing regulating riboswitch. The design was made to begin the branch point at the end of random region of the library. The 5' primer binding site of this library will include the 5' splice site. This last construct was chosen, because of the "ready-to-screen" design. Indeed, the enriched pool could be directly cloned in the pWHE601* plasmid (116) and candidates that regulate gene expression can be spotted by *in vivo* screening method described in the chapter 3.1.5.

The first test was to check if the construct of the library with this splice site as a docking sequence could be splice out in yeast without inhibition of gene expression. For this, the plasmid pWHE601* and the library were double digested by Age-I HF and Nhe-I HF. These fragments were afterwards ligated and transformed in *Escherichia coli* Top 10 competent cells. One of the colonies formed was used to inoculate a 4 mL overnight culture in LB-Amp media. The cells were harvested and their plasmids recovered using a Qiagen Plasmid Miniprep Kit. The plasmid sequence was checked by sanger sequencing (Seqlab laboratories) transformed into yeast competent cells and spread on a SCD-URA plate. Three different colonies were picked and separately used to inoculate 4 mL liquid

culture and incubated overnight at 30°C and 150 rpm. Cells were washed two times in 10 ml PBS and fluorescence was measure with the Fluorolog. The result shows that there is no decrease in fluorescence expression with the insertion of a random sequence as an intron compared to the positive control expressing constitutively the GFP (figure 3.5.). In fact, the expression level is even slightly higher with the random pool, due to the expression enhancement due to splicing.

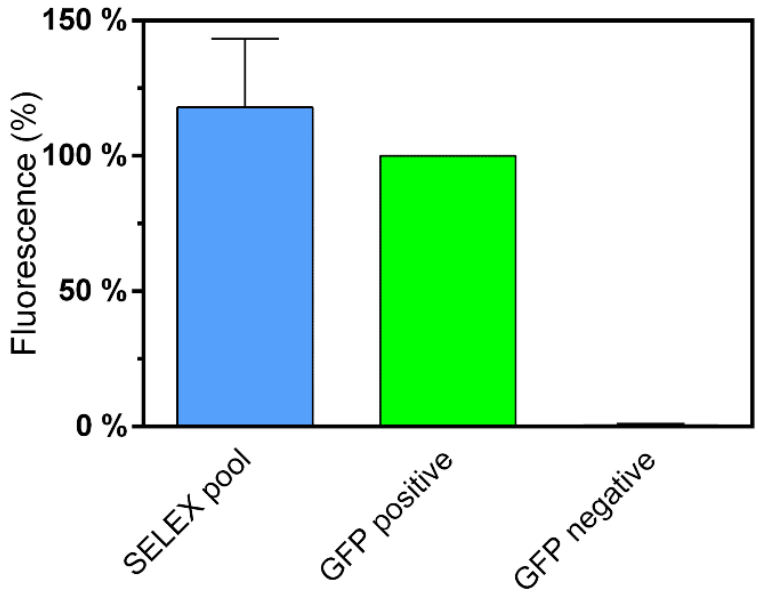

Figure 3.5. Fluorescence measurement of the pool when inserted in the pWHE601* plasmid. After insertion of the library in the GFP gene, the fluorescence of the transformed yeast was measured (blue bar) to verify if this construct was still allowing gene expression. This fluorescence was compared to yeast constitutively expressing the GFP (green bar). The negative control (no GFP expression) is represented on the right of the graph).

3.1.1.2. Design of the CO sequence.

Now that the design of library was chosen, the sequence and the condition of the hybridization need to be established. The first design of the CO was made to include the branch point region (6 nts) and to that was added the 6 following nucleotides in order to have a melting temperature of the capture oligo superior at the room temperature (in this case Tm = 30°C). The sequence was 5'-TAC TAA CTT CTA - 3'. The interaction CO-RNA library was checked by EMSA (Electro Mobility Shift Assay) with a 15% native polyacrylamide Gel run at 4°C. The CO was radioactively labelled, mixed with the RNA library, folded under different conditions and the shift of the band was compared between each

experiments. Each best result was kept for the following test. The CO concentration was fixed to 10 µM.

- In PBS with a 1:5 ratio CO/RNA:
  - No folding
  - 95°C 5’ and ice 5’
  - 95°C 5’ RT 5’
- In PBS with a 1:5 ratio CO/RNA, 95°C 5’ and at RT for:
  - 20’
  - 2h
  - Overnight
- In PBS with a 1:5 ratio CO/RNA, 65°C 5’ and RT 20’ plus:
  - 0 mM Mg
  - 5 mM Mg
  - 10 mM Mg
  - 15 mM Mg
- In PBS + 5 mM magnesium with a 1:5 ratio CO/RNA, 65°C 5’ and RT 20’ plus:
  - 0 mM NaCl
  - 200 mM NaCl
  - 400 mM NaCl
  - 600 mM NaCl
- In PBS + 5 mM magnesium, 65°C 5’ and RT 20’:
  - 1:5 CO/RNA ratio
  - 1:10 CO/RNA ratio
  - 1:20 CO/RNA ratio
  - 1:30 CO/RNA ratio
  - 1:50 CO/RNA ratio
  - 1:120 CO/RNA ratio
  - 1:180 CO/RNA ratio

After all these experiments, only the ratio CO/RNA and the presence of magnesium seems to have a positive impact on the amount of complex formed. However, the amount of CO shifted on the gel is really low in all these experiments, even with a high ratio RNA:CO (figure 3.6.A.).

As a last option, the CO was extended with one guanosine on the 5’ side to increase the Tm of 2 degrees and form a stronger bond. As shown on figure 3.6.B. this experiment (done with a 1:5 CO:RNA ratio and a folding protocol of 5’ at 65°C and 20’ at RT) led to a dramatic increase in the CO shift. Now that the hybridization CO:RNA was proven, the formation of the complex beads-CO-RNA needed to be tested.

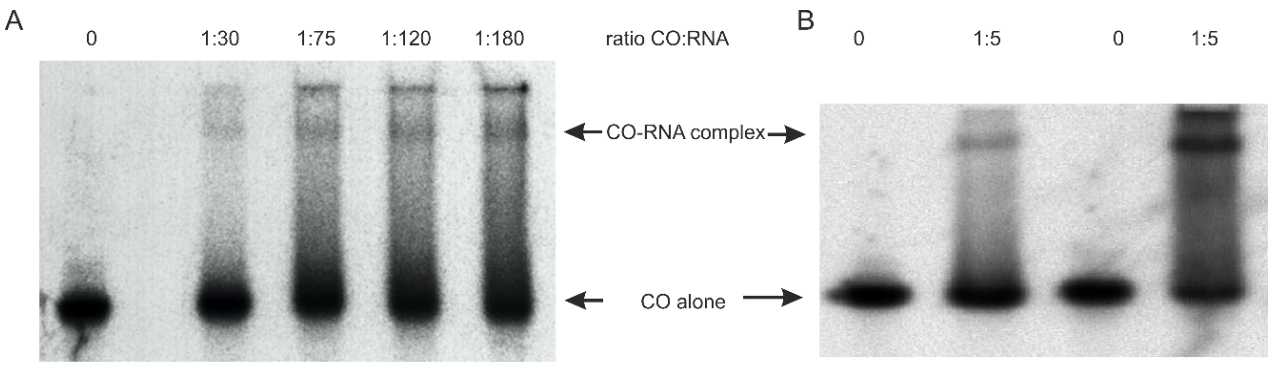

Figure 3.6. EMSA gels for CO-RNA immobilization verification. (A) EMSA gel between the CO and a gradient of concentration of RNA. The CO was radioactively labelled and folded with either no RNA, 1:30, 1:75, 1:120 or 1:180 CO:RNA ratio and loaded on a 15% native polyacrylamide gel (from left to right on the gel). The CO alone and the complex CO-RNA are annotated with arrows. The higher the amount of RNA, the more CO is shifted. However, even in the best case, the amount of CO shifted represent only a small fraction of total amount of CO. (B) EMSA gel of the RNA library comparing the CO length. The two gel lines on the left were done with the 12 nt long CO and the two gel lines on the right were done with a 13 nt long CO. Both experiment were done without RNA pool (-) and with RNA pool (+). A clear shift can be observed for the 13 nt long CO, unlike for the 12 nt long where the shift is very weak.

Choice of the matrix and of the CO linker

The CO is the complementary sequence of the DS to which a linker has to be added to be able to have an anchor to a matrix. The first approach tried was to add a primary amino group linked through a C12 long aliphatic chain to the 5' side of the CO and to use epoxy activated sepharose to form a covalent bond between the amino residue of the linker and the matrix. To verify the immobilization efficiency, an immobilization test was realized comparing the CO with linker to the CO without linker. In this experiment, 1 mL of sepharose powder was transferred into a falcon tube, filled up to 5 mL with DMSO. Then, 10 nmol of CO with or without ligand was mixed with the sepharose and incubated for 4 h using a revolver shaker at room temperature. The sepharose was afterwards washed with 5 mL DMSO and incubated with 0,5 mL of 0,1 M ethanolamine to saturate all the reactive site still available. Both column material was rinsed again with 5 mL DMSO and transfer in column. RNA was meanwhile prepared by performing a body-labelling transcription using alpha-^{32}P-UTP in the *in vitro* transcription (detailed in table 5.4.). RNA was purified by ammonium acetate precipitation and solved in water. The radioactivity of the transcript was measured and 500.000 counts was boiled to 95°C 5 min, then directly transfer to ice for 5 min and finally mixed with 0,1 V of 10X PBS and incubated 15 min at room temperature. The whole solution was transferred inside of the columns containing 0,5 mL of sepharose derivatized either with the CO with linker or the CO

without linker. The flow-through liquid was collected and the column was further incubated at room temperature for 15 min, 30 min, 45 min and overnight. After this incubation time, columns were washed 4 times with 0,5 mL PBS. The washing fraction were recovered and 10 µL were transferred into scintillation vials containing 2 mL of scintillation liquid for radioactivity measurement. 10 µL of the column material was also harvested and transfer into the scintillation vials to measure the relative amount of RNA hybridized on the CO in the column.

After comparing the amount of radioactivity in the column indicating the immobilization rate of the RNA library, with and without a linker on the CO, no difference was observed at different incubation times, indicating that the CO is not able to anneal with the library (figure 3.7.).

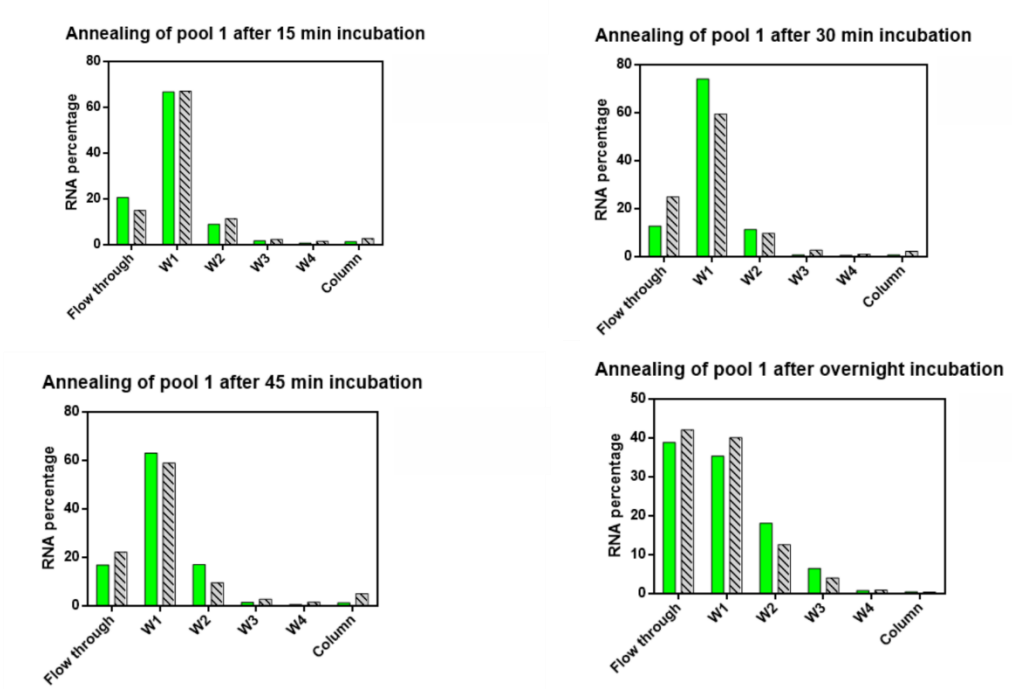

Figure 3.7. Comparison of the immobilization rate of the pool on the CO with and without linker. Both CO with and without linker were separately mixed with an epoxy-activated sepharose. Only the CO with linker should be able to immobilize and to retain the library in the column. Radioactively labelled RNA was loaded onto the column material with CO+linker (green) and without linker (grey). This mixture was incubated 15, 30, 45 min and overnight before starting the washing. In all cases, after 4 washes, no radioactivity was detectable on the column.

An explanation may be that the epoxy activated column seems also to react with the primary amines of the bases on the CO and create *ipso facto* covalent bond in a non-desire position. To circumvent this, the approach was changed to a biotin-streptavidin interaction between the CO and the matrix because of its specificity and the removal of any type of chemistry in the process. Indeed, such a technique requires only an incubation of streptavidin and biotin in physiological condition during

some minutes to form interaction almost as stable as a covalent bond. The detailed immobilization process will be described in 3.1.1.1.5.

Establishment of the folding protocol

The folding step is the first step of the SELEX procedure. Folding the pool starts in heating the pool until a specific temperature is reached where all the library is single stranded and denatured and then to cool it down. This step is necessary because of the properties of single stranded nucleic acid. Indeed, DNA or RNA which is not double stranded can fold in several different secondary structures. Some of these structure can be folding traps in which the aptamers cannot be able to bind the target. In the original protocol, the DNA is folded by a heating step at 90°C for 8 min and immediately cooled down at 4°C for 10 min followed by a short incubation at room temperature. In parallel, streptavidin-coated magnetic beads (10^9 beads in the first round and 10^8 beads in each of the following rounds) is washed 3 times with 500 µL B&W buffer (binding and washing buffer, 10 mM Tris-HCl pH 7.5, 1 mM EDTA, and 2 M NaCl). The beads are separated by placing the tube in a magnet stand. After washing the beads, they are resuspended in B&W buffer to a concentration of 2×10^9 beads/mL, and an equal volume of the CO is added (600 pmol biotinylated oligo/1×10^8 beads). They are mixed together for one hour with gentle rotation. Then the DNA is added to the already perform complex CO-Magnetic beads.

However, a modification in the protocol was introduced. As it is already known, a library of random single stranded nucleic acid can fold in many different structures because of the huge diversity of the pool. If the folding of the RNA is done before, in most of the cases the DS will be hidden and not available for hybridization with the CO. The DS will be in most case base pair internally with a part of the random region (especially if the library is undergoing an important and fast decrease of the temperature, some sequences can be trapped in the closest local minimum energy) and will not be able to overcome this energy when put in contact with a short oligonucleotide. Based on the primer annealing principle for the hybridization of short primer, the folding of the RNA was done in presence of the CO with a slow cooling down to room temperature to force most of the library to bind to the CO. As the folding was done with SELEX buffer in which magnesium was present for the folding of the aptamer, but also to enhance the interaction DS-CO and the interaction RNA-Target, the folding protocol can't go to 95°C without cleaving the RNA. Therefore, and based on EMSA experiments, the folding protocol consists in mixing CO and RNA in 1X PBS supplemented with 600 µM NaCl and 5 mM magnesium, heat it up to 65 °C for 5 min to unfold the RNA without degrading it and then 20 min at 21°C to stabilize the complex CO-RNA.

Assembly of the complex Beads-CO-RNA

The first step of the process is to incubate the radioactively labelled RNA here in excess (ratio CO:RNA > 1:5) with the CO (600 pmol) in hybridization buffer (1x PBS, 600 µM NaCl, 5 mM $MgCl_2$). The mixture was incubated 5' at 65°C to unfold the RNA without cleaving it and transferred into a thermoblock already precooled at 21°C for 20'. The complex RNA:CO was afterwards mixed with 150 µL of beads (10 mg/mL) and incubated 1 h at RT. The beads were previously washed 3 times with B&W buffer (10 mM Tris, 500 mM NaCl, 20 mM EDTA) according to the manufacturer instruction. After the incubation, the beads were pelleted with a magnetic rack and the supernatant removed. The wash fractions were transferred in scintillation vials for radioactivity measurement to quantify the relative abundance of RNA in each fraction. The beads were resuspended in 200 µL SELEX buffer (40 mM HEPES, 250 mM KCl, 20 mM NaCl and 5 mM $MgCl_2$). This volume was chosen because it's big enough to wash efficiently the beads with efficiency. The volume can be pipetted at once with the 200 µL pipette to avoid accidental carry on of magnetic beads while the aspiration of the liquid. The beads were washed 10 times with 200 µL of SELEX buffer and were transferred to a scintillation vials to measure the amount of RNA still immobilized on the beads. A control experiment was performed in parallel without CO to determine the amount of non-specific interaction of the RNA with the beads (figure 3.8.). As observed on the graph, a clear specific binding among the three parties can be observed. Indeed, 20% of the total radioactivity was detected on the beads in the experiment with the CO. The negative control without CO shows less than 1% of RNA sticking to the beads. The data shows also that after three washes, the amount of RNA removed by washing can't go any lower indicating that all RNA not immobilized on the CO are removed. With this last result, the assembling of the complex was successful and the next step can be analysed.

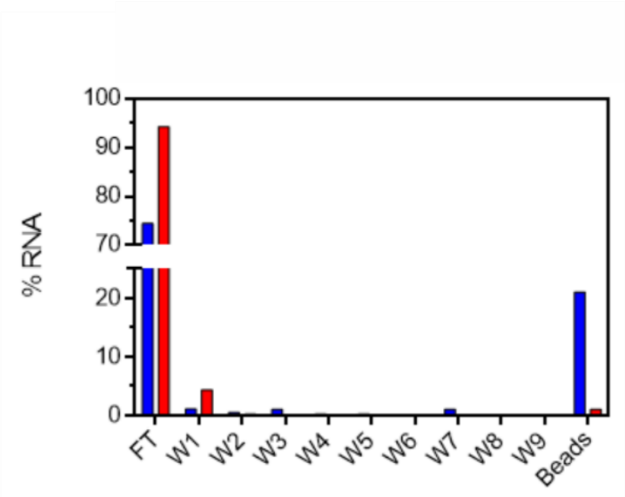

Figure 3.8. Test of the assembly Beads-CO-pool. The protocol of immobilization was performed on a mixture with CO (blue) and without CO (red). After two washes, the background is already reached for both experiment. However, the experiment with the CO shows a remain of 20 % of the RNA on the beads after 10 washes. The graph on the right is a zoom in the left graph to clearly observe the value of the wash steps.

3.1.1.2. Pool production and reamplification step

Now that the design of the pool and CO is ready (figure 3.9), and the assembly of the complex beads-CO-pool are established, the RNA library needs to be produced. The DNA template needs to be order in sufficiently high amount to cover a theoretical diversity of 10^{15}-10^{16} different sequences (between 2-20 nmol). Furthermore, the optimal PCR conditions and PCR efficiency need to be determined to generate the biggest DNA template possible for later *in vitro* transcription. Finally, testing the RT-PCR efficiency with this library.

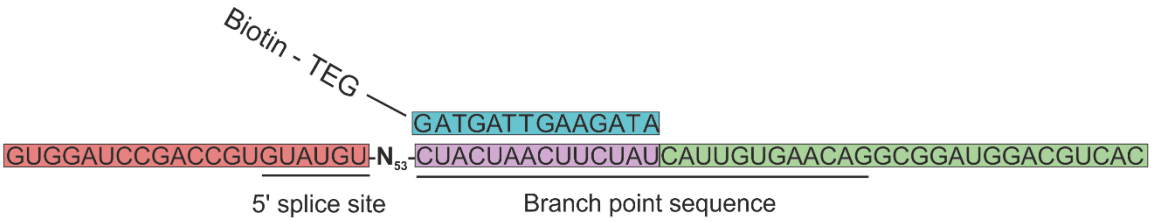

Figure 3.9. Design of the Capture-SELEX library for splicing regulation. The 6 last nucleotides of the 5' primer binding site (red box) contains the 5' splice-site. The N53 are located just afterwards. The docking sequence (purple box) starts with a C and followed directly by the branch point and the 3' splice site sequence until the G101. The 3' primer binding site is in green box. The CO (blue box) is coupled on its 3' side to a TEG linker terminated by a biotin to allow immobilization on the streptavidin-coated-beads.

Establishment of the PCR conditions

The first test was classical template amplification with Taq polymerase. The PCR conditions are listed in material and methods. The PCR mix was prepared and launched for 15 cycles. At round 5/7/9/11/13/15 10 µL of the reaction mix were took out and stored on ice. A 3% agarose gel was made to determine the number of cycles where overamplification starts to appear. Overamplification needs to be avoided, especially during the amplification of the original template. Indeed, because of its randomized content, some random region of a sequence can form a duplex with the random part of another sequence and amplify wrong sequences that can be transcribed afterwards but will be totally different as the original template. No overamplification was observed at round 5 and a little bit at round 7. Therefore, annealing temperature optimization test was done with 4/5/6/7 cycles of PCR at three different annealing temperatures 54°C/58°C/62°C. As observed on the gel (figure 3.10.), the optimal annealing temperature is 58°C. There is no difference seen between 54 °C and 58 °C, but to enhance specific amplification of the pool and to decrease the risk of size product amplification, the highest annealing temperature was chosen.

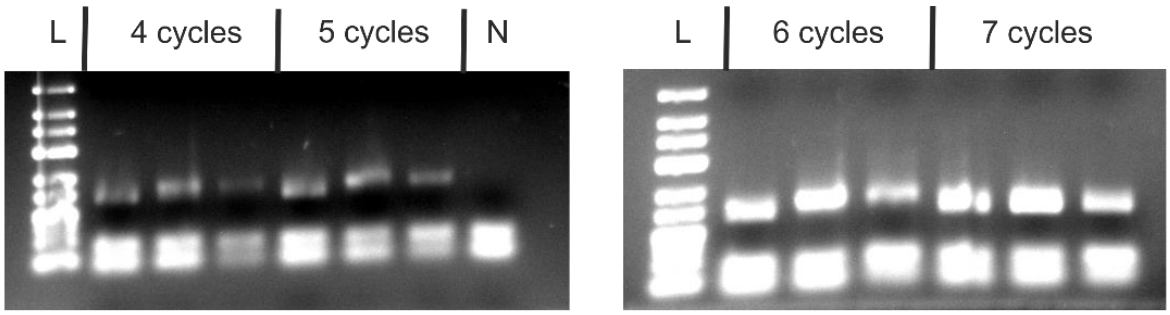

Figure 3.10. Annealing temperature optimization test. The test was performed with 4, 5, 6 and 7 cycles at three different annealing temperatures: 54°C (left band), 58°C (middle band) and 62°C (right band). The strongest band can be observed at 7 cycles for both 54°C and 58°C. L represent the ladder and N a negative control without template.

Finally, the PCR efficiency was determined. A good value needs to be the closest to 100% if the PCR conditions are optimized. A too much different value will indicate bad primer design or non-optimal reagent concentration or reaction conditions. A serial dilution of the template was realized from 2 µM to 1.95 nM with 11 steps of 2-fold dilution in between. 10 µL of each concentration was added to the PCR mix and 5 cycles with 58 °C annealing temperature PCR was performed and 10 µL was afterwards loaded on a 3% agarose gel (figure 3.11.). The PCR efficiency was calculated using ImageJ by calculating the intensity of each band on the gel. The intensity was plotted in function of the

template concentration and the slope value was reported. This value was entered in the thermofischer PCR efficiency calculator and was determined at 103% indicating a high PCR efficiency.

Once the selection step is finished, the eluted RNA needs to be synthesise into cDNA and then amplify by PCR. This step is the reverse-transcription PCR (RT-PCR). Therefore, a RT-PCR test of the pool needs to be perform to ensure that this step is working. Therefore, RNA was *in vitro* transcribed from the DNA template and purified with a standard ethanol sodium acetate precipitation. The RT-PCR mix was prepared as seen on table 5.9 (see Methods). 1 pmol of RNA and in 50 µL SELEX buffer was added to two different reaction tubes containing the RT-PCR mix. A 65°C unfolding step for 5 min was performed and then 1 µL of each enzyme (SSII and Taq pol (NEB)) were added to the mix. 10 min at 54 °C for the RT reaction and 6 cycles of PCR were enough to see a clear band on the gel for 1 pmol of initial RNA template which correspond the expected result.

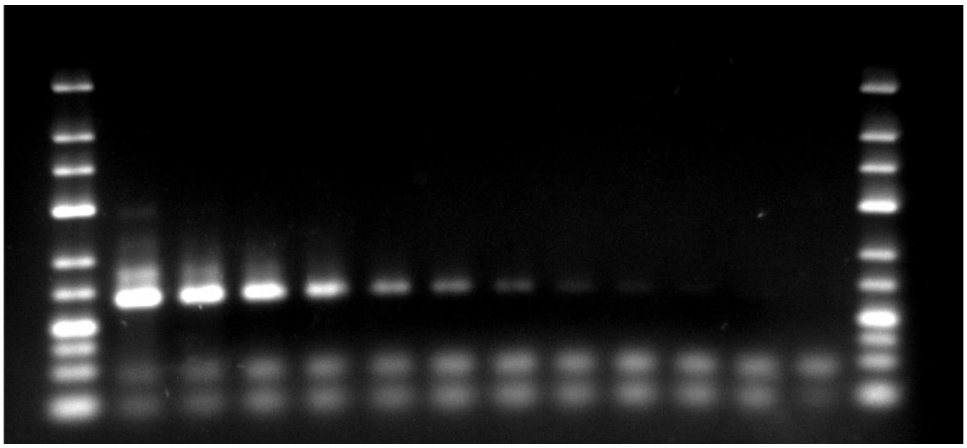

Figure 3.11. PCR efficiency test. Test PCR was performed with 58°C annealing temperature and 5 cycles. 12 reactions were performed in parallel with a decreasing concentration of template (from left to right 2 µM / 1 µM / 500 nM / 250 nM / 125 nM / 62.5 nM / 31.25 nM / 15.63 nM / 7.81 nM / 3.9 nM / 1.95 nM).

RNA library production

Based on the optimal PCR conditions determined before, 10 mL of PCR reaction based on the previously mentioned protocol. The mix was made with 10 nmol of pool-template oligonucleotide based on the design of the library chosen in 3.1.1.1.1. The mixture was aliquoted at 100 µL in a 96 well plate. 5 cycles of PCR with an annealing temperature of 58°C was performed. DNA product was afterwards pooled together and purified by sodium acetate precipitation and solved into 2 mL of milli-Q water. Once purified, a 100 µL test transcription was performed using 2 µL of the DNA template to check the size of the transcription product. A control gel of this *in vitro* transcription was performed (figure 3.12.).

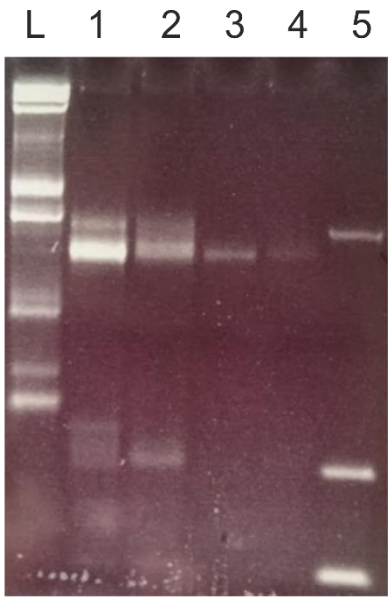

Figure 3.12. Control gel of the in vitro transcription. After a 1h test transcription, 5/2/1/0.5 µL were loaded on a 6 % denaturing polyacrylamide gel (respectively line 1, 2, 3 and 4 from left to right). The first line on the left is the low range RNA-Ladder (peqGOLD). The line 5 on the right was loaded with the PCR template only. The size of the transcript corresponds at the size of the library (123 nt).

The control gel showed that the transcription was producing the expected size RNA. Therefore, a larger scale transcription was performed in order to generate enough RNA for several selections. The total volume of the transcription was fixed on 10 mL to ensure the production of enough RNA. The composition of the transcription mix is detailed in the material and methods. The amount of PCR template added in the transcription reaction has to be chosen in order to have at least a theoretical diversity of the library of $5*10^{15}$ different sequences. The PCR was performed with 10 nmol of template oligonucleotide, equivalent to $6,2*10^{15}$ different sequences. 5 cycles of PCR were performed, meaning that 16 copies of each sequence were created making about 10^{17} sequences of DNA template in the 2 mL of stock generated. Therefore, one tenth of the DNA template was taken out for the transcription, containing 10^{16} sequences, which are big enough to cover entirely the maximum diversity of the sequences from the original template oligonucleotide. The RNA was transcribed overnight at 37°C and precipitated afterwards with a sodium acetate precipitation. Once purified, the RNA was loaded on 4 different 6% PAA gels to purify it and recover only the desired size RNA to avoid contamination by parasites. RNA is cut out from the gel, the band is chopped on transfer into a falcon tube containing 10 mL of NaAc 300 mM per gel performed and incubated overnight at 4°C with mild overhead shaking to make the RNA diffuse out of the gel. Gel pieces are discarded and the liquid is filtrated to remove small pieces of gel and precipitated by the addition of

2,5 volumes of a 1:1 mixture of ethanol/acetone. RNA is afterwards resuspended in 1 mL of milli-Q water and its concentration is determined with a nanodrop measurement.

3.1.1.3. Proof of principle, SELEX against neomycin (NEO)

The selection process must happen in a buffer system where ions concentration and pH are defined. As we are aiming for *in vivo* active riboswitch in yeast, the potassium and sodium concentration in the buffer were designed to mimic the cytosolic conditions of yeast. Magnesium concentration is however critical for aptamer folding and binding. However the magnesium concentration in yeast is relatively low (between 0,1 and 1,0 mM (117)) compared to the concentration most often used in SELEX (between 1 and 10 mM). To ensure that no good potential binders will get lost during selection because of low magnesium concentration, 5 mM will be used in the SELEX buffer. Following buffer was used as SELEX buffer: HEPES 40 mM (pH=7,4), KCl 250 mM, NaCl 20 mM and $MgCl_2$ 5 mM.

As all the step of SELEX are established and tested, a first test selection can be performed to see if an enriched pool can be created with our design. The target chosen is neomycin B. An aptamer against NEO was already found by classical SELEX (72) and NEO has characteristics to facilitate interaction with RNA (6 amino groups and it's a ribosome targeting drug). Hence it makes an ideal proof of principle target. As seen on the figure 3.8., 3 washes of 200 µL after the flow through are enough to reach the plateau of remove non-specifically RNA and weak binders. As a theoretical 2 nmol of purified RNA, equivalent to a diversity of 10^{15} different sequences were used for the first round. According to the manufacturer, 1 mL of dynabeads have the capacity to bind 2 nmol of biotinylated ss-oligonucleotides which correspond to the amount of RNA used for the first round. For all the following rounds ten times less beads will be used for the selection given that less RNA need to be screened.

The first round selection happens as follow. 1 mL of magnetic beads is transferred to a 1.5 mL Eppendorf tube and placed 1 min on a magnetic rack. The supernatant is removed and the beads are washed with 500 µL of B&W buffer according to the manufacturer instructions. Then, the beads are washed again with 500 µL of SELEX buffer and resuspended in 200 µL of SELEX buffer. Meanwhile, the 2 nmol of RNA, 2 nmol of CO are mixed in 200 µL of hybridization buffer, heated up to 65°C for 5 min and then transferred to a thermoblock pre-set at 21°C for 20 min. Afterwards the RNA-CO complex and the beads are mixed with the beads and incubated at room temperature under upside-down shaking using a revolver mixer for one hour. After the 1h immobilization the supernatant is removed (flow-through) and stored in an Eppendorf tube. The beads are resuspended in 400 µL of

SELEX buffer and incubated 5 min on the revolver shaker. The supernatant is again removed and the wash step is repeated 3 more times (four times 5 min wash in total). The last wash step was performed for 30 min in SELEX buffer to keep the same conditions as in the elution step. After the wash steps, 400 µL of 1 mM neomycin B in SELEX buffer is added to the beads and incubated again 30 min on the revolver shaker. The supernatant is afterwards recovered and mixed with 480 µL of 2x RT-PCR buffer, heated up to 65°C for 5 min. After the heat step, 10 µL of SSII and 10 µL of Taq Pol are added to the mix. The mix is afterwards split in 10 PCR tubes of 100 µL and a 4 cycles RT-PCR program is launched (10 min at 54°C; 4 cycles of {20 sec 95°C; 20 sec 58°C; 20 sec 72°C}). A 3% agarose gel is afterwards performed on the PCR product to check for correct amplification of the product and to control the absence of over-amplification. Then 10 µL is used again for an overnight body labelling with alpha-^{32}P-UTP transcription The transcription is afterwards purified by ammonium acetate precipitation to get rid of the radioactive nucleotide and solved in 50 µL water. From this purified RNA, 1 µL is measured to check the amount of count per microliter. Then, 500.000 counts are taken for the next round.

The protocol for all the rounds after the first round is slightly adapted to smaller volumes because of the use of only 150 µL of beads instead of 1 mL. There are only 2 wash steps of 200 µL in SELEX buffer. The last wash and the elution with the target are performed in 50 µL. This allows to use solely one RT-PCR mix that can be directly transferred to the elution fraction and then start the heating step at 65°C for 5min for the RT-PCR protocol. The number of PCR cycles is adjusted in function of the presence or the intensity of the band on the gel. It can change between 6 and 14 cycles depending of the stringency.

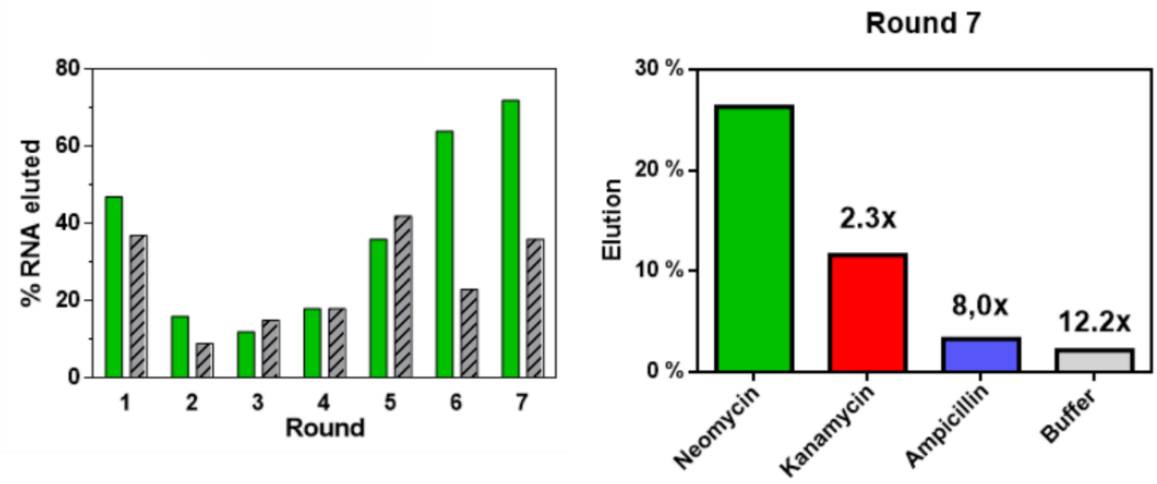

Figure 3.13. Capture-SELEX against NEO. The green bars represent the amount of RNA specifically eluted from the beads by addition of 1 mM of NEO (left graph). The grey bars represent the amount of RNA eluted at the last wash step. The enrichment is visible when there is at least twice the amount of RNA eluted in the specific elution than in the background elution. This phenomenon is visible at round 6 already. The specificity at the round 7 was tested by making in parallel 4 experiment with 4 different conditions for elution (right graph). The bars represent the amount or RNA eluted by each condition. KANA eluted 2.3 times less RNA than NEO. AMP used as a random target for negative control eluted as much RNA as only buffer.

After 6 rounds of selection, enrichment could be observed in the elution fraction (figure 3.13. left). The enrichment can be detected when the amount of radioactivity measured in the elution fraction is two times or higher superior at the amount of radioactivity in the last wash fraction. A clear enrichment against NEO is observed in this case. To test if the enrichment was specific and not due to enrichment of weak CO binders, the round 7 was performed with 4 reactions in parallel: One of them with NEO, the other with kanamycin (another aminoglycoside), the third with amoxicillin (another random antibiotic) and with only buffer. As observed in the graphic (figure 3.13. right), the library is specifically enriched towards NEO. A fraction of the pool is able to bind also kanamycin but the amount is more than two times lower than for NEO. This is due to the similarities between the two molecules. Amoxicillin is not able to elute more RNA than only buffer. With this first experiment, we proved that a fast and specific method for RNA aptamer selection where the ligand is free in solution was successfully developed. We also proved that the design of our library is suitable to find aptamers using Capture-SELEX.

3.1.2. Capture-SELEX against new targets, a mixture of paromomycin, amoxicillin and citrate (PAC)

An advantage of the Capture-SELEX strategy that was not mentioned before is the possibility of introducing a mixture of several targets at the same time in the elution step. By using this approach, we can theoretically select for three targets in one tube as long as they are not interacting with each other. As a first try, a mixture of paromomycin (PARO), amoxicillin (AMOX) and citrate (CIT) was used for the next SELEX. Each target was at a final concentration of 1 mM in the elution solution. To ensure that they do not have negative influence on each other, the three antibiotics were incubated together and separately in SELEX buffer in the freezer overnight and was thawed to check if there was the formation of any precipitate or any change of colour with a fresh prepared mix. As the three ligands seems to not interact with each other, the SELEX can be performed with this mixture. The selection was performed in the same conditions as before.

The same library used for the NEO SELEX was used. Again, 2 nmol of the RNA pool were used as a starting pool in the first round of selection. A radiolabelled tracer was incorporated inside the pool to estimate the quantities of RNA present in every step of the selection. The first round was performed like for the NEO SELEX first round and all the other round were performed like the round 2 of the NEO SELEX. For all the rounds, RNA was eluted with 1 mM of each target mixed in SELEX buffer to be sure to elute most of the binders. After 9 rounds enrichment was observed with a factor of 2.9-fold difference between the background elution on the specific elution. This enrichment was confirmed at round 10 with a 3.7-fold difference.

The following step was to determine which target(s) was responsible for the enrichment. For this, the eleventh round of SELEX was performed in 3 different reaction tubes, except that the elution solution was containing only one of the targets (one different for each tube). As it can be observed on the figure 3.14., only PARO was responsible for the library enrichment (4.14-fold elution). The two other targets did not have a bigger impact on the elution of RNA compare to SELEX buffer meaning that the library enriched only for one target, in this case PARO.

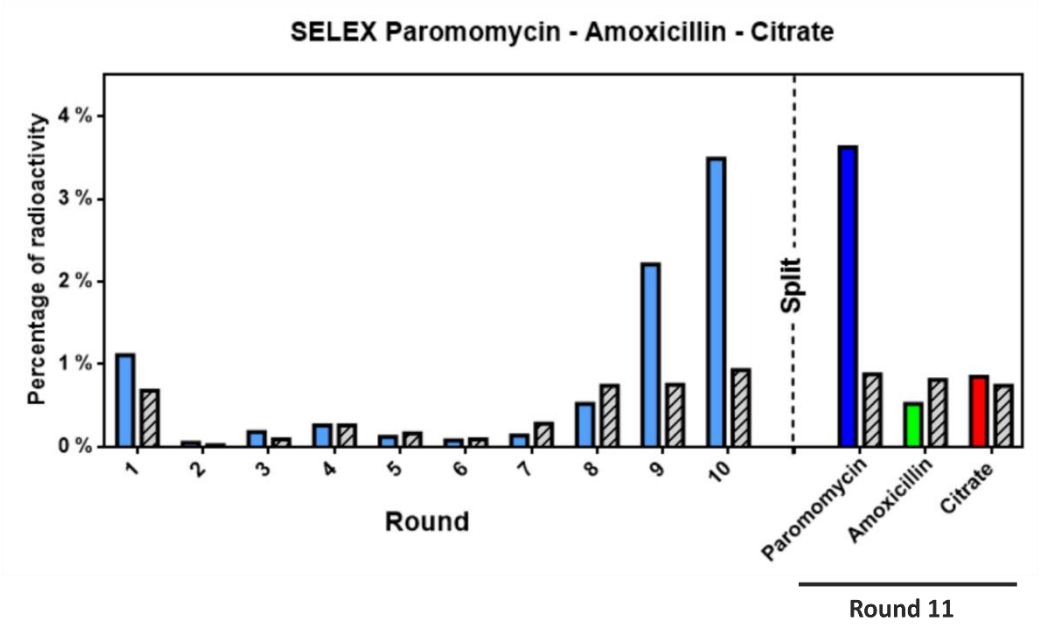

Figure 3.14. Capture SELEX against PAC mixture. The light blue bars represent the amount of RNA eluted with a mixture of 1 mM of PARO, AMOX and CIT each. The round 11 was performed with each target separately with PARO (dark blue), AMOX (green) and CIT (red). Only PARO was able to enrich the library. The two other compounds weren't able to undock more RNA than only buffer.

3.1.3. *In vivo* screening for splicing regulating aptamers

The underlying idea of this library design (mentioned in 3.1.1.1.1) was to be able to generate splicing regulating riboswitch directly out of the SELEX procedure. To test this hypothesis, the round 9 and round 11 of the PAC SELEX were cloned into the pWHE601* plasmid in the 5'UTR of the GFP gene. Each round was amplified again by PCR using a Q5 polymerase (polymerase with a proof reading domain, making less mistakes than the Taq) with 10 rounds to avoid overamplification. The DNA was afterwards purified, digested in 50 µL with both Age-I HF and Nhe-I HF enzyme and again purify to remove the small digested fragment at the extremities. The pWHE601* plasmid was digested by the same enzymes and purified on a 1% agarose gel to eliminate non-digested plasmid. The backbone and the PCR product of both rounds were assembled by T4 ligase ligation, purified by butanol precipitation and transformed in competent *E.coli* cells. From the outgrowth, 1, 10 and 100 µL were plated on LB-Amp plates to estimate the size of the library transformed. The rest from the 1 mL outgrowth step were used to inoculate a 4 mL LB-Amp liquid media. On the plate inoculated with 1 µL 52 colonies grew, on the plate with 10 µL plate around 500 colonies were estimated and the plate with 100 µL too much cells to be counted grew. The amount of different sequences cloned in the plasmid is estimated at about 50.000 which is diverse enough to perform screening. Cells from the liquid culture were harvested and their plasmid recovered. 100 ng of plasmid were used to transform 50 µL competent yeast cells and were plate on 20 SCD-ura plates to allow the growth of only cells

that integrated the plasmid. After 48 h of incubation at 30 °C, between 200-300 yeast colonies grew per plate. Colonies were picked and were transferred into a 96 well plates (1 colony per well, 200 µL of SCD-URA media per well) and transferred in a 30 °C incubator shaking at 1200 rpm. After 24 h, 20 µL of each well were transferred in another 96 well plate containing 180 µL of SCD-URA with either 100 µM of PARO or no PARO. After screening 1092 different colonies of each round, no candidate showing change in fluorescence were observed, hence no candidates that regulate gene expression with splicing using PARO was found. After observation of the results, the design for splicing regulation was thought as too complicated to hope finding a candidate regulating splicing. Indeed, based on the tetracycline riboswitch that regulates splicing, a precise optimization of the genetic context was important which is not possible with the randomized library. Therefore, the library has to be changed for being suitable for translation regulation, where the regulation is less dependent on capturing sequences but where only the stability of the complex formed can regulate gene expression. Furthermore, this strategy permitted the discovery of the neomycin, tetracycline and ciprofloxacin riboswitch in their original state.

3.2. Optimisation of the protocol and selection against a different range of target

After the success of the two previously mentioned selections (NEO and PARO), the next step was to enrich more libraries to prove the efficiency of this new protocol. Therefore, several selections were performed against mixtures of different targets. One of them was a penicillin antibiotic mixture (penicillin G, penicillin V, ampicillin, amoxicillin and 6-aminopenicillanic acid) and the other one was a mixture of metabolites (fructose-bi-phosphate, phosphoenol pyruvate and D-ribose-5-phosphate). Unfortunately, none of the selection showed an enrichment after 15 rounds. However, the experiment was done in the same conditions as the previous one. These results pointed that the protocol need some improvements to find aptamers for other targets that are more difficult.

Indeed, some molecules, due to their properties, are more predisposed to interact with nucleic acids. RNA is negatively charged, is highly soluble in water, can make hydrogen bonds with the Watson-Crick or Hoogsteen interaction side, and contains rings that can allow stacking interaction. With all these information, we can determine if a target is “easy” or not. A target such as neomycin is positively charged, made of sugar rings, highly soluble in water, has a lot of moieties able to do hydrogen bonds and is a known antibiotic targeting the ribosome which is made of RNA. All these characteristics make neomycin an “easy” target for selection and therefore justify the choice of it for the establishment of the protocol.

3.2.1. New pool design for ribosome scanning blocking

The first design of the pool for splicing control had led to no success in the screening for *in vivo* riboswitches. Therefore, a new pool was designed (shown in figure 3.15.) ready to clone in the 5'UTR of the GFP gene. This pool was designed for finding riboswitch that can block the scanning of the ribosome upon ligand binding. First, the primer binding site were designed based on the pWHE601* plasmid (used for *in* vivo screening) to allow direct homologous recombination by the primer binding site are already included in the backbone. For the 5' primer binding site, the sequence was elongated by the T7 promotor to allow *in vitro* transcription. Then three Gs were attached at the 5'end to facilitate transcription *in vitro* using T7 RNA polymerase and afterwards of 18 nt that precede the Age-I restriction site on the pWHE601* and finally of the Age-I restriction site. The 3' primer binding site contains first the "Kozak-start sequence" followed by the Nhe-I restriction site and the 21 nt downstream the restriction site on the backbone.

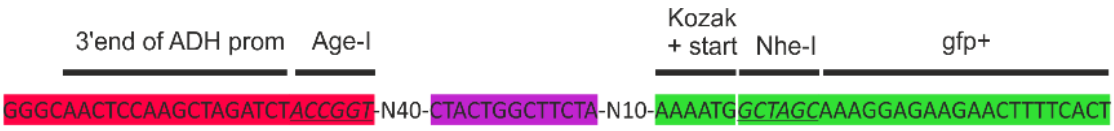

Figure 3.15. Design of the new Capture-SELEX library. The 5' (red) and 3' (green) primer binding site are made based on the sequence located around the two restriction sites Age I and Nhe I in the screening vector (italic and underlined) to allow using of homologous recombination and to minimize the impact of the backbone on the riboswitch. The docking sequence (purple) was mutated on the 6th and the 7th nucleotide by two G.

As the first design of the docking sequence allowed the successful enrichment of the NEO and PARO SELEX, no major change was made on the design. However, the splicing region had to be mutated to inhibit any interaction with the spliceosome during the later *in vivo* screening process. Furthermore, as we already observed before, the higher is the Tm of the complex CO-DS, the more RNA will be immobilized on the beads. Consequently, the nt A6 and A7 were replaced by two guanosines to enhance the stability of the complex. Moreover, with this design, the DS is partially complementary to the 5' primer binding site. Hereby, we wanted to enhance the switching capability of the generated aptamers by giving a refolding possibility to the docking sequence when they are release from the CO upon ligand binding. Finally, the DS design has to avoid any internal structure to be formed such as stems to enhance the binding to the CO by not having the DS folded on itself.

Unlike the first design, the choice was made to introduce a short N_{10} random region on the 3' side of the DS. We observed that this region exercises a selection pressure on the random region on the

5’ side of the DS. Indeed, after enrichment of the PARO SELEX, all the most enriched sequence contained a fully complementary sequence to the sequence located on the 3’ side of the DS, thus generating a very long and stable stem loop that might block completely the ribosome scanning even without target binding.

3.2.2. Optimization of the time of elution

In the current protocol, a time of 30 min for elution was chosen. The theory was that the longer the elution time is, the more binders will be recovered. However, based on the measurement of the wash 3 (W3) compared to the wash 2 (W2) (W2 is 5 min and W3 is 30 min), the amount of RNA in the W3 is significantly higher than in the W2. This indicates that a long time of washing may lead to more unspecific undocking. To test this hypothesis, the enriched round 9 in the PAC SELEX was used in 4 experiences in parallel, where different times (5, 15, 30, 120 min) were used for W3 and the specific elution (SE) (figure 3.16).

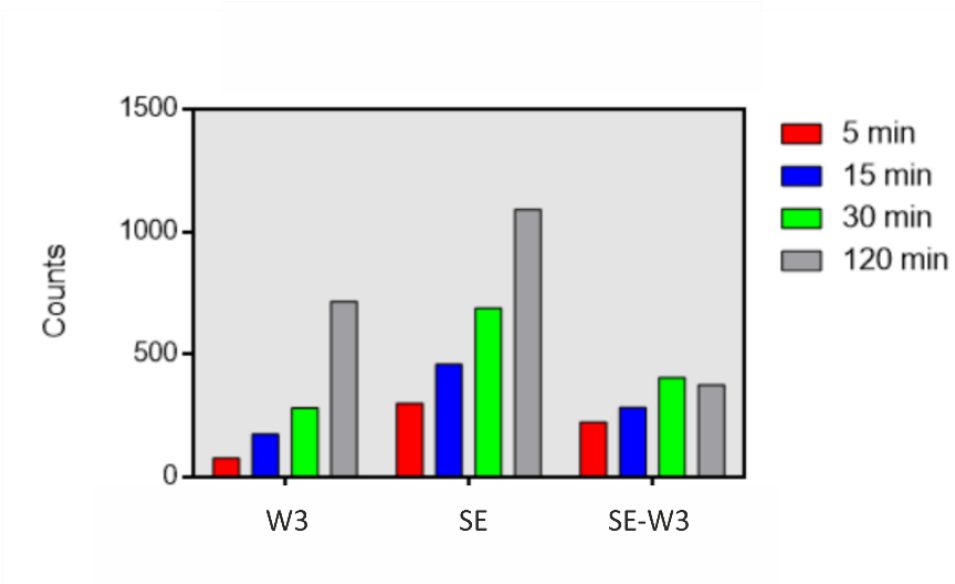

Figure 3.16. Comparison of the incubation time of the last washing step (W3) and the specific elution (SE) on the round 9 of the PAC SELEX. The bars represent the amount of RNA detected in the W3 fraction (left) and in the specific elution (SE) fraction (centre) eluted with 1 mM PARO for 5 (red), 15 (blue), 30 (green) and 120 min (grey). The difference between the SE and W3 (SE-W3) is displayed on the right to compare the efficiency in elution. In each case, the longer the wash is, the more RNA is recovered but in both non-specific and specific elution. However, for 120 min washing the difference between SE and W3 is lower than for 30 min, indicating that a long washing time makes only undock more unspecific RNA.

The result shows that the longer the time is the more RNA gets eluted in both fraction. Though, the ratio of SE/W3 which can be interpreted as specific elution on unspecific undocking gets smaller when the time is longer. This strongly suggests that a long time of washing is not beneficial for the

selection because instead of privileging the aptamers elution, a longer time of elution will only bring more easy-undockable sequence. Hence, the elution time for W3 and SE steps were changed to 5 min like all the other washing steps in order to improve the partitioning effect.

3.2.3. *In vitro* selection using the optimized Capture-SELEX pool design and protocol

3.2.3.1. PARO SELEX

First, the improvements of the pool and protocol should be tested against a target that already worked with the previous design, therefore PARO was used as a target. If the improvements were really efficient, an enrichment should be observable at the same round of selection or even earlier. In addition, we wanted to develop a PARO specific riboswitch that can discriminate with NEO. Indeed, there is already a NEO riboswitch that does not respond to PARO. Therefore, we aimed for a riboswitch that is able to switch with PARO but not NEO.

The new library, prepared in the same conditions as mentioned before, containing 10^{15} different sequences (=2 nmol of RNA), was used as starting pool for the first round of selection. A radiolabelled tracer was incorporated to estimate the quantities of RNA present in every step of the selection. In the first round 1 mL (=10 mg) of magnetic beads with a binding capacity of 2 nmol of single-strand DNA were prepared, like mentioned in the first chapter, and used to be able for the immobilization of the RNA. 150 µL of magnetic beads were used for all the other cycles. The first seven rounds RNA was eluted with 1 mM of PARO in SELEX buffer to be sure to elute most of the binders. After the confirmed enrichment of the pool towards PARO in round 6 (figure 3.17.), the concentration of PARO was decreased to 100 µM (rounds 8-9). With this strategy, only aptamers with a K_D sufficiently low will be able to undock from the capture oligonucleotide, bind to the target, and stay on the target. Increasing the stringency in this way will pressure the pool to select the aptamers with the best affinities. A counter-elution step with 100 µM NEO was introduced in the two last rounds (round 10 and 11) of selection not only to increase affinity but also to enrich for PARO binding aptamers that do not recognize the closely related molecule NEO. After two rounds of counter elution, we observed already that the amount of RNA eluted with PARO was two times higher than the one which are eluted in the counter-elution (figure 3.17.). The detailed summary of the selection can be found in table 1.

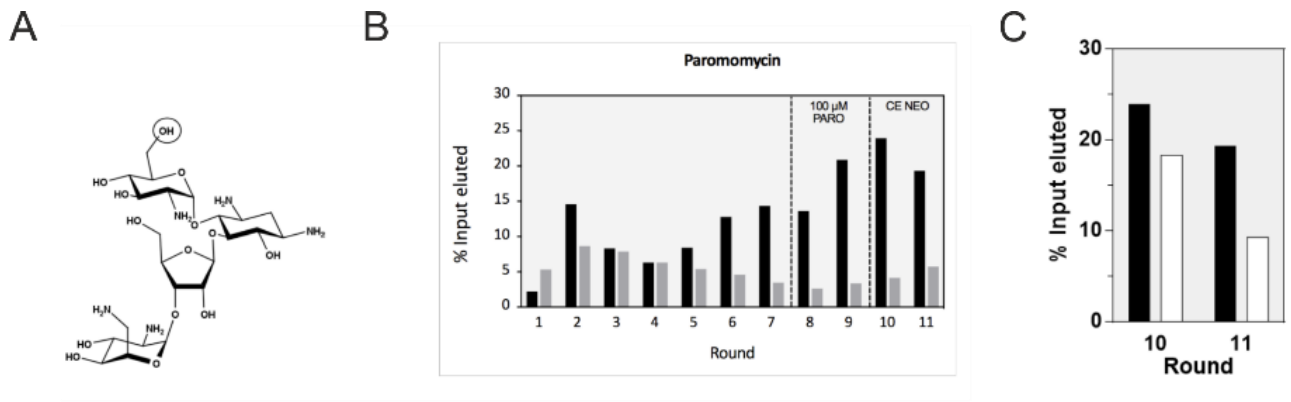

Figure 3.17. Capture-SELEX against PARO. (A) Chemical structure of the PARO on the left. PARO and NEO are very structurally similar. They differ only from a hydroxyl residue (PARO) to an amino residue (NEO) on the carbon 6 of the first hexose. (B) SELEX results. The black bars represent the amount of RNA specifically eluted from the beads by addition of PARO. The grey bars represent the amount of RNA eluted at the last wash step. The dashed lines represent an increasing of the stringency. To create a specific pool to PARO, a counter selection with NEO was performed at round 10 and 11. (C) After two rounds, the amount of RNA eluted with 100 µM NEO (white) is two times lower than with 100 µM PARO (black).

Table 1: Detailed summary PARO SELEX

Round	Eluent	Counter-elution	Number of wash	Ratio SE/W3[1]
1	1 mM PARO	-	3	0,41
2	1 mM PARO	-	3	1,69
3	1 mM PARO	-	3	1,06
4	1 mM PARO	-	3	1,00
5	1 mM PARO	-	3	1,56
6	1 mM PARO	-	3	2,78
7	1 mM PARO	-	3	4,14
8	0.1 mM PARO	-	3	5,20
9	0.1 mM PARO	-	3	6,19
10	0.1 mM PARO	0.1 mM NEO	3	5,76
11	0.1 mM PARO	0.1 mM NEO	3	3,38

[1] SE: Specific Elution, W3: wash 3

The new SELEX against PARO was successful, not only because of the enrichment of the library in an earlier round, but also because an *in vitro* selectivity was applied with success by enriching RNA that only bind PARO but not NEO.

3.2.3.2. SELEX against kanamycin B and viomycin

The PARO selection was successful and an enrichment could be detected. However, other selections need to be perform to show that the selection strategy is working. Therefore, other compounds that are known to interact with RNA were tried. Kanamycin and viomycin are both antibiotics that act on the bacterial ribosome despite that they don't belong to same class and should be also successful for the SELEX. Selection for both target was made in parallel, and an enrichment was observed for both SELEX (figure 3.18.).

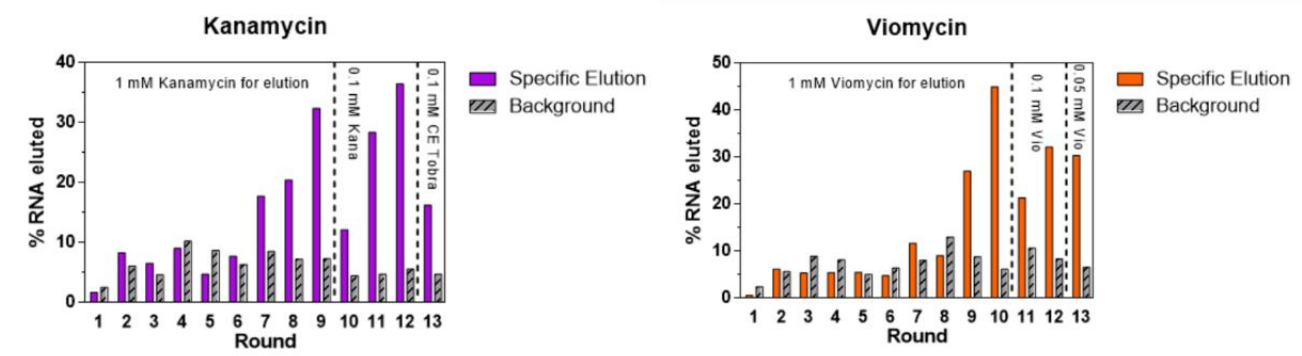

Figure 3.18. Capture-SELEX against kanamycin (left) and viomycin (right). The coloured bars represent the amount of RNA specifically eluted from the beads by addition of the target. The grey bars represent the amount of RNA eluted at the last wash step. The dashed lines represent an increasing of the stringency. The concentration of the targets was reduced to 0.1 mM at round 10 for KANA and 11 for VIO. The concentration for VIO was even reduced to 0.05 mM at the last round. To create a specific pool to KANA a counter selection with TOBRA was performed at round 12 and 13.

All selections steps were performed with the same conditions (number and volume of washing step, CO and beads quantities…) as for the PARO SELEX. For kanamycin 13 rounds were performed, two rounds of SELEX with 100 µM elution was performed (rounds 10 and 11) plus two rounds with counter elution with 100 µM tobramycin (round 12 and 13). For viomycin, 13 rounds were also performed, elution was decreased for two rounds (round 11 and 12) at 100 µM and for the last round at 50 µM, indicating a theoretical low K_D for this target.

3.2.3.3. SELEX against penicillin derivative antibiotics and metabolites

As the selection with the previous design against these targets did not worked, they were tried again with the new design. Furthermore, these targets are more difficult to find an aptamer against compared to the previous ones, they were mixed together to enhance the probability of finding a new aptamer. The selection was done using a mixture containing 1 mM of amoxicillin, ampicillin, penicillin G, penicillin V and 6-aminopenicillanic for elution. For the metabolites, fructose

bisphosphate, phosphoenol pyruvate and D-ribose 5' phosphate was mixed together. Unfortunately, with the current protocol both of selections didn't shows an enrichment pattern after 15 rounds (data not shown). Interestingly, the number of PCR cycles performed after the selection never went above 6 rounds, indicating a high number of RNA eluted, even in the early round where only few RNA are supposed to be undock. Therefore, the SELEX was performed a second time with a loading of only 100.000 counts of RNA on the beads instead of 500.000 counts. Hereby, we wanted to decrease the RNA concentration on the beads to avoid the unspecific undocking of too much RNA during the specific elution step. The strategy was successful for decreasing the amount of RNA undocked from the beads. Indeed, the number of PCR cycles increased from 6 cycles to an average of 10 cycles for the first 10 rounds, with a maximum of 16 cycles for the second round. However, this did not lead to a visible enrichment meaning that no aptamer was enriched during this selection.

3.2.3.4. SELEX against ATP and chloramphenicol

ATP is a nucleic acid analogue that can perform Watson-Crick interactions and chloramphenicol is an antibiotic targeting the ribosome and possesses two chlorine moieties that are very electronegative. According to their characteristic, both targets are adapted target to find aptamer against. Hence, a SELEX was performed in the same conditions as before. However, after 10 rounds, no sign of clear enrichment was observed despite the predictions made based on our experience. Interestingly, a general increase in the amount of RNA on the beads and eluted at each wash was observed, suggesting a non-specific enrichment.

3.2.2. Aiming for a better partitioning

Improvement for a better partitioning was targeted in different ways. Indeed, there are several possibilities to decrease background elution in the Capture-SELEX. The following possibilities were tested: reduction of the carried-over beads, washing strength and the length of the CO.

3.2.2.1. Reduction of the carried-over beads

One of the technical difficulties in the manipulation of beads is that they stick on the tube wall or on the pipet tip. This leads sometimes to the transfer of RNA loaded beads to the specific elution fraction that get amplified afterwards in the PCR step. To decrease this effect, Tween 20 was added to the washing buffer at a concentration of 0,002%. At this concentration, the beads get all grouped together in one point when the magnet is applied and there is no effect on any of the other steps of the SELEX.

Another simplification step was tried in the protocol. In the current protocol the hybridization buffer (1x PBS, 600 µM NaCl, 5 mM $MgCl_2$) used is different than the SELEX buffer (HEPES 40 mM pH = 7.4, KCl 250 mM, NaCl 20 mM, $MgCl_2$ 5 mM). A quick test was performed to check if the immobilization was identical in SELEX buffer. Two simultaneous SELEX rounds were performed with the two buffers to compare the result and the impact on the immobilization was compared. As expected, no difference was observed in both experience. Therefore, the SELEX buffer will also be use in the future as immobilization buffer.

3.2.2.2. Washing strength

As seen before, the partitioning efficiency can be impacted by the washing steps. Indeed, at a fixed stringency, after an infinite number of wash almost all RNA will be theoretically removed from the beads. Furthermore, not all sequences behave like one unique sequence with their interaction with the CO. Some of them can be qualified as "weak binders", meaning that because of their structure, not all the docking sequence is able to interact with the CO or they can change their folding when a little bit of energy is brought to the beads (shaking or pipetting). All of these weak sequences have a different strength of attachment to the CO. So, we could apply to the beads one step with a higher washing stringency. Consequently, all the weakest binders will get remove from the beads. When the selection conditions will be changed back to softer conditions, all the sequences left will be average and strong binders. These sequences will tend to stay more on the beads instead of being washed away during the specific elution step. This should lead to elution only of real binders during the specific elution step.

One of the approach tried was to use a heating step during the washing of the RNA. A higher temperature will remove all the sequences that have a weak binding affinity to the CO. For optimization, the SELEX of ATP done previously was used. Indeed, in this SELEX the amount of RNA eluted in both W3 and SE fraction was getting higher and higher. The explanation could be that an enrichment of specific binders was happening but as too much background RNA was present, the enrichment could not be detected. The round 11 was performed identically as round 10, except for the introduction of the heating step during wash 1. The beads were resuspended in 200 µL SELEX buffer, but instead of placing them on the revolver shaker, the beads were heated up to 309°C for 5 min. The amount of RNA removed in the two first wash steps went from 37% to 48% in the new protocol. On the other hand, the specific elution fraction contained much less RNA. This modified round was performed again 4 more rounds until a clear enrichment was observed (figure 3.19. on the left). Three more rounds were performed with ATP concentration of 0.1 mM.

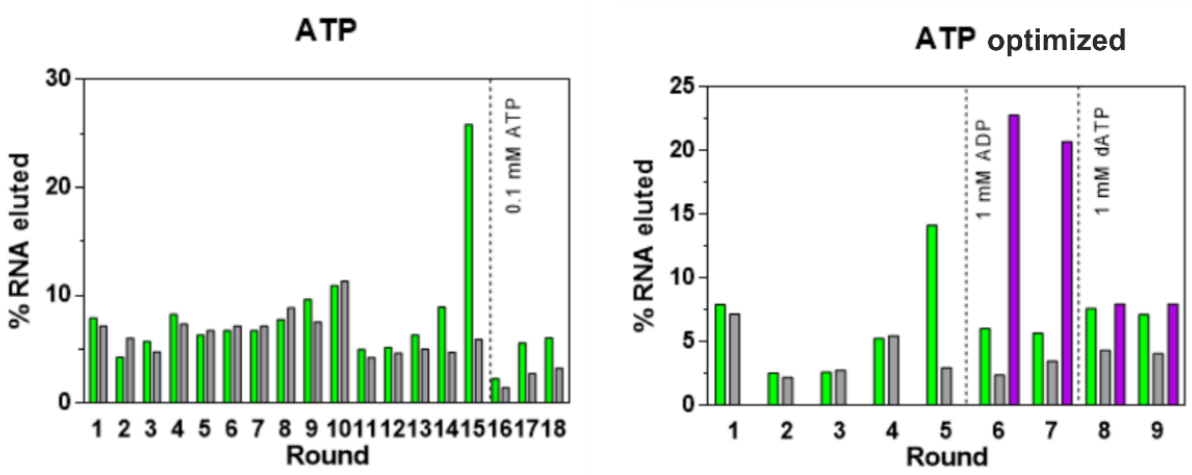

Figure 3.19. Capture-SELEX against ATP. The green bars represent the amount of RNA specifically eluted from the beads by addition of the target. The grey bars represent the amount of RNA eluted at the last wash step. The dashed lines represent an increasing of the stringency. The purple bars represent the amount of RNA in the counter-elution. For the non-optimized protocol (left), the concentration of the target was reduced to 0.1 mM at round 16 after the confirmed enrichment for three more rounds. For the optimized protocol (right), counter elution with 1 mM ADP was performed directly after the round 5 for two rounds and a second counter-elution phase was performed with dATP for the two last rounds.

To check if the enrichment could be detected earlier, the RT-PCR fraction from the round 1 was reused for body labelling transcription and the round 2 of SELEX was redone but with the heating step. As observed in the figure 3.19. right, the enrichment can be clearly detected at round 5. This SELEX was used afterwards to create a ATP specific pool by the use of counter-elution with 1 mM ADP and 1 mM dATP. Unfortunately, almost all the enriched sequences were also able to bind ADP with the same affinity than ATP. Indeed, the amount of RNA removed during the counter-elution steps, never decrease and the specific elution was almost as low as the background. However, the enriched pool seems to discriminate better dATP to ATP than ADP to ATP despite that dATP only lack an oxygen atom on the ribose and that ADP has a complete phosphate group missing (figure 3.20.).

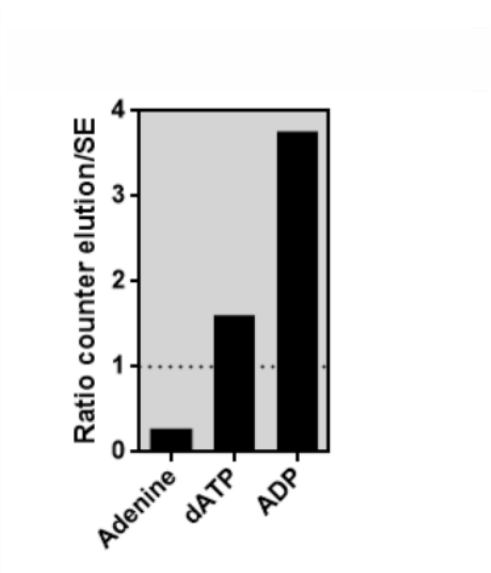

Figure 3.20. Specificity of the round 6 of the ATP SELEX. The bars represent the ratio between the amount of RNA eluted in the counter elution (CE) step for adenine, dATP and ADP (respectively from left to right), A value above 1 indicates that more RNA was eluted during the CE step than for the SE. The higher the bar, the more this difference is.

In order to verify if this was not an isolated case, this heat step improvement was applied to the SELEX of kanamycin and viomycin but from the very beginning. If any clear difference is observed, we can conclude that the heating step was really beneficial to find more enriched pool (figure 3.21). For the kanamycin SELEX, the enrichment could be observed at the round 5 already, which is two rounds earlier than the previous one. Two step of decreasing of target elution to 100 µM was applied and two rounds of counter-elution with tobramycin was also added afterwards. The efficiency of the counter-elution was tested afterwards for the round 6 (first enrichment), round 9 (before CE) and for round 11 (after counter elution).

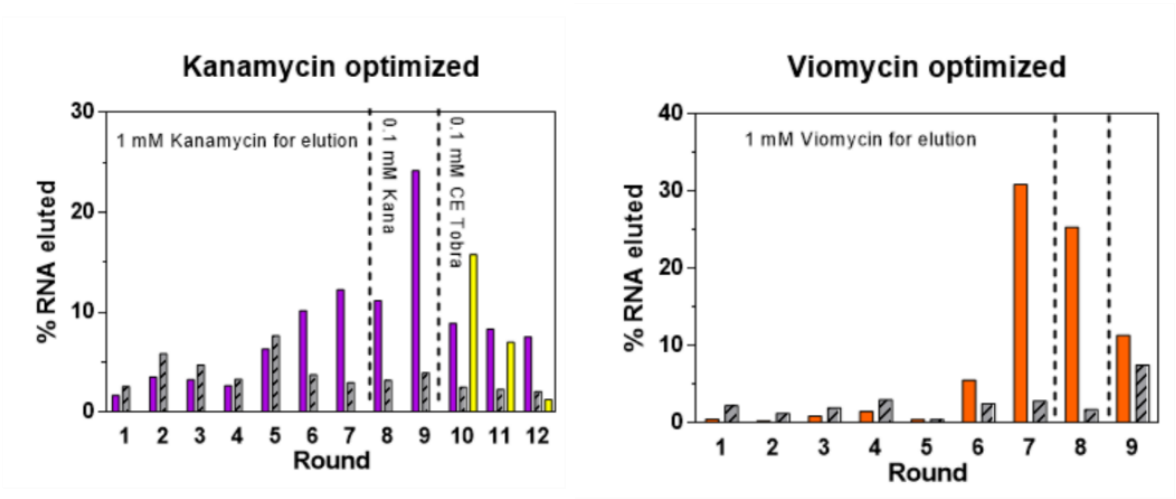

Figure 3.21. Optimized Capture-SELEX against KANA and VIO. The purple and orange bars represent the amount of RNA specifically eluted from the beads by addition of respectively 1 mM KANA and VIO. The grey bars represent the amount of RNA eluted at the last wash step. The dashed lines represent an increasing of the stringency. The yellow bars represent the amount of RNA in the counter-elution eluted with 1 mM TOBRA. In both cases, enrichment can be detected at round 6.

The data showed (figure 3.22.) that at the first enrichment, only KANA is able to elute RNA. However, after three more rounds of SELEX, a significant part of the pool is able to bind TOBRA. These binders are afterwards removed from the SELEX thanks to the CE step, because not more RNA than in the W3 can be detect in the CE fraction. PARO was also tried in parallel at each round but was never able to elute RNA. For the viomycin SELEX, the enrichment was observed at round 6, which is 3 rounds earlier than the previous SELEX. Concentration of viomycin was decrease to 100 µM at round 8 and 9 to achieve a better K_D of the library.

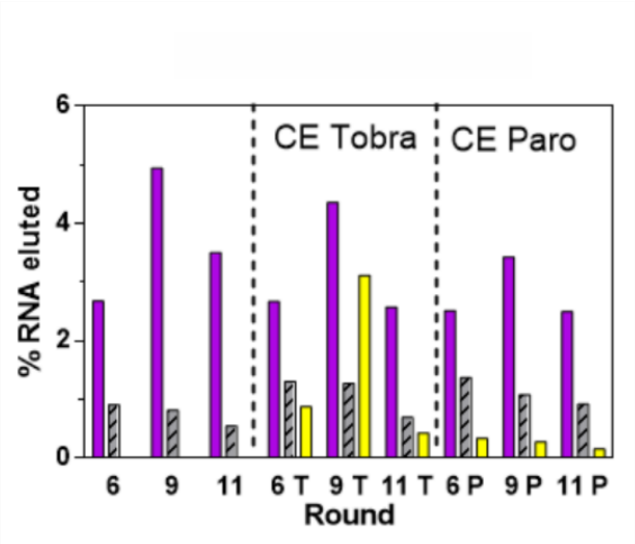

Figure 3.22. Specificity test on the optimized kanamycin SELEX. The round 6, 9 and 11 were used for this test and treated like in a classical SELEX round. Three experiments were done in parallel, one without counter elution (left), one with counter elution with tobramycin (center) and one with PARO. The purple bars represent the amount of RNA eluted by kanamycin the grey bars are the RNA detected in the W3 and the yellow bars are the RNA detected in the counter elution.

3.2.2.3. Capture-oligonucleotide length

The last aspect studied in this thesis was the variation of the length of the CO. As seen during the establishment of the method, a longer CO involves a higher melting temperature, meaning a bigger stability of the complex RNA-CO. However, the design of the docking sequence is fixed since the beginning of the SELEX and can't be changed during the selection. Hence, an approach with a modified CO was tried. Three different CO were designed with extensions. But as the bases around the docking sequence are random, the modified CO were ordered with extension of 1, 2 and 3 random nucleotides on each side. If we consider that an RNA will not be attracted to its exact CO extension match, meaning the binding to the CO will be independent of the extended nucleotides of the CO, we can estimate the probabilities of matching to the table 2.

Table 2: Calculation of the matching base pairing possibilities with CO extension

Type of CO	Probability of no binding to the extended region	Exactly one binding to the extended region	Exactly two binding to the extended region
CO +2 (N-CTACTGGCTTCTA-N)	56%	37%	6%
CO +4 (NN-CTACTGGCTTCTA-NN)	32%	42%	21%
CO +6 (NNN-CTACTGGCTTCTA-NNN)	18%	35%	32%

Observing this statistics, these modified CO will behave mostly like an extension of one or two nucleotide of the CO and should not lock the aptamer too tightly to the CO. A test experiment was done using the P11.2_H2 aptamer from the PARO SELEX and using the enriched round 9 from the KANA SELEX. If the extension of the CO is working, we should observe a decrease in the last W3 and in the SE fraction, due to the decrease of non-specific RNA undocking. The results from the P11.2_H2 candidate shows logically a decrease in the relative amount of RNA eluted from the beads when the CO became longer. However, the result from the enriched round 9 from the KANA SELEX doesn't show any clear pattern, indicating that a longer CO is may be not better than the original one but has however no bad influence (figure 3.23).

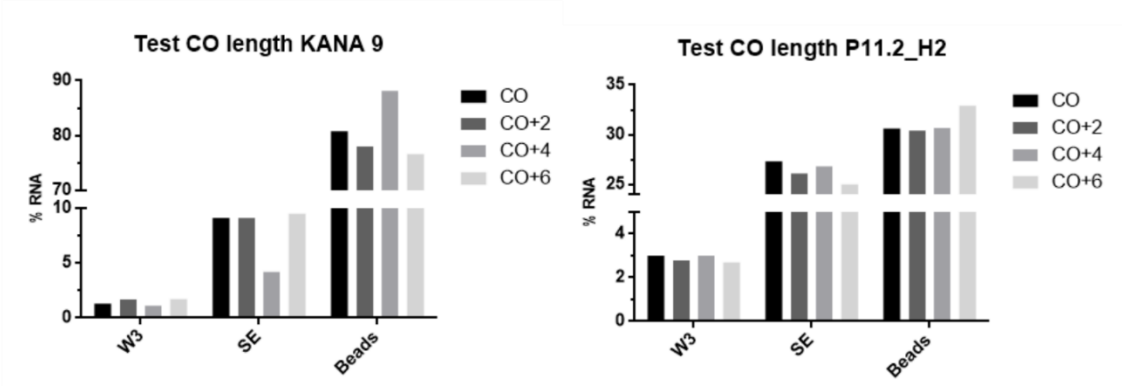

Figure 3.23. Test length CO on the KANA 9 pool (left) and on the P11.2_H2 (right). The amount of RNA found in the fraction W3 (left), in the SE (centre) and on the beads (right) is represented for the 4 different lengths of CO: original (black), CO with 1N on each side (dark grey), CO with 2N on each side (grey) and CO with 3N on each side (light grey). For the KANA 9 CO +4, no visible pattern can be observed. For the P11.2_H2, RNA in the SE and in the W3 slightly decreased when the CO length increased. The amount of RNA on the beads was as expected higher with a significant longer length of CO.

3.2.3. Selections with the optimized protocol

A combination of all the improvement suggested before was tried to check if they theory all together could select aptamer on harder targets. To this purpose, a SELEX was done again a mixture of target that didn't worked before and some other hard target of interest (chloramphenicol (CMP), fructose bisphosphate (FBP), glyphosate (GLY), strep tag and IPTG). The SELEX was done with the inclusion of

the previous optimization and along with the elongation of the CO during the SELEX and a very fast elution time. The exact protocol is described in table 3.

Table 3: description of the optimized protocol for the mixture selection.

Round	Time of elution	CO length	Heating step
1	30 min	original	-
2	10 min	CO +2	-
3	5 min	CO +2	28 °C
4	5 min	CO +4	28 °C
5	5 min	CO +4	28 °C
6	5 min	CO +6	30 °C
7	10 sec	CO +6	30 °C
8	10 sec	CO +6	30 °C
9	10 sec	CO +6	32 °C
10	10 sec	CO +6	32 °C
11	10 sec	CO +6	32 °C

The SELEX results are shown in figure 3.24. An enrichment could be detected at round 10 and confirmed at round 11.

As the SELEX was made with a mixture of targets, a deconvolution was performed to find which ligand(s) enriched the library (Figure 2-L). Interestingly, the pool was not enriched against one ligand but slightly against two ligands that made together the enrichment. The successful targets were glyphosate and FBP. Glyphosate slightly eluted more RNA than FBP, suggesting that there are more glyphosate aptamers in the pool than FBP aptamers. Two rounds more were performed separately with only GLY or FBP to eliminate the binders to the other target. Unfortunately, no visible enrichment was observed for FBP.

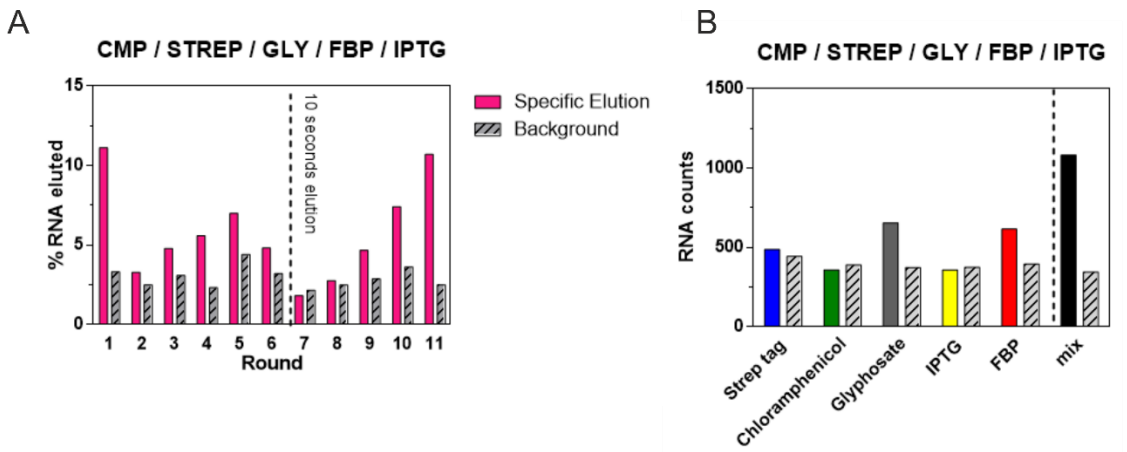

Figure 3.24. Capture-SELEX against a mixture of ligand (CMP, STREP, GLY, FBP IPTG). (A) The pink bars represent the amount of RNA specifically eluted from the beads by addition of the mixture. The grey bars represent the amount of RNA eluted at the last wash step. The dashed lines represent an increasing of the stringency by reducing the elution time to 10 seconds. The value of the SE in the first round is very high due to the SE time of 30 min compare to the W3 time of only 3 min. Enrichment can be visible at round 10. (B) Binding of the pool from the mixture to each ligand separately. The round 11 of SELEX was performed with each target separately and the amount of RNA eluted was compared. The coloured bars represent the amount of RNA eluted with the target, the bar in black represents the amount of RNA eluted by the mixture of all. The grey bars represent the amount of RNA in the fraction W3. Only GLY and FBP are able to elute more RNA than the background, meaning that they co-enriched together the library.

3.2.4. Development of a C-less SELEX

The high flexibility of the aptamers is making them good for binding targets by adopting a broad range of different conformations. The most stable interaction intramolecular in the nucleic acid is the GC base pairing. Theoretically, by removing the possibilities of GC base pairing formation, the flexibility of the RNA could be increase by allowing the formation of only AU base pair or GU wobble base pair. Therefore, a new pool adapted for Capture-SELEX was designed for this. The random region was order with only A, G or U possibilities. In the docking sequence, the first and last C is exchange to a G and the primer binding sites are kept identical. The SELEX was performed in the same conditions as the KANA and VIO optimized SELEX. A selection was performed against neomycin as a control. However, no enrichment was visible after 8 rounds indicating that this design is unable to enrich specific aptamers against even a good target such as neomycin.

3.3. Development of a synthetic paromomycin riboswitch

3.3.1. *In vivo* screening for searching a paromomycin riboswitch

In our screening system, aptamers that can regulate gene expression upon ligand addition of the media were looked for. In this case, aptamers that can act as a roadblock to the 30 S subunit ribosome are detectable. For this purpose, the SELEX library is cloned in front of a fluorescent reporter gene in a plasmid. Yeast cells are afterwards transformed with this plasmid and plated. Single colonies are picked and transfer to a preculture. This preculture is afterwards split in two different wells, one without target and one with target. After letting them grow until they reach the stationary phase, cell fluorescence is measured and compared. If the fluorescence of a well without target is significantly higher than its equivalent incubated with target (at least 1.5-fold), then the sequence inserted in front of the reporter gene is able to fold in a more stable structure upon target binding. Consequently, the 30 S subunit of the ribosome scanning process is inhibited by its roadblock due to the aptamer increased stability.

RNA libraries from round 8 and 11 were cloned via homologous recombination into pCBB06 (described in the figure 3.25) in the 5' UTR of a GFP expressing gene. The pools were integrated in yeast using homologous recombination and analysed to see fluorescence change upon paromomycin addition. A first step of sorting using FACS were performed to remove all the cells not expressing GFP which are cells either with a too strong RNA structure cloned in front of the reporter gene or only the empty backbone. For the round 8, 80% of the cells were GFP positive and for the round 11 only 50% (figure 3.26). In total, more than 2.000.000 cells were collected. This effect is due to the selection process that selects more and more structured aptamer during the selection experiment.

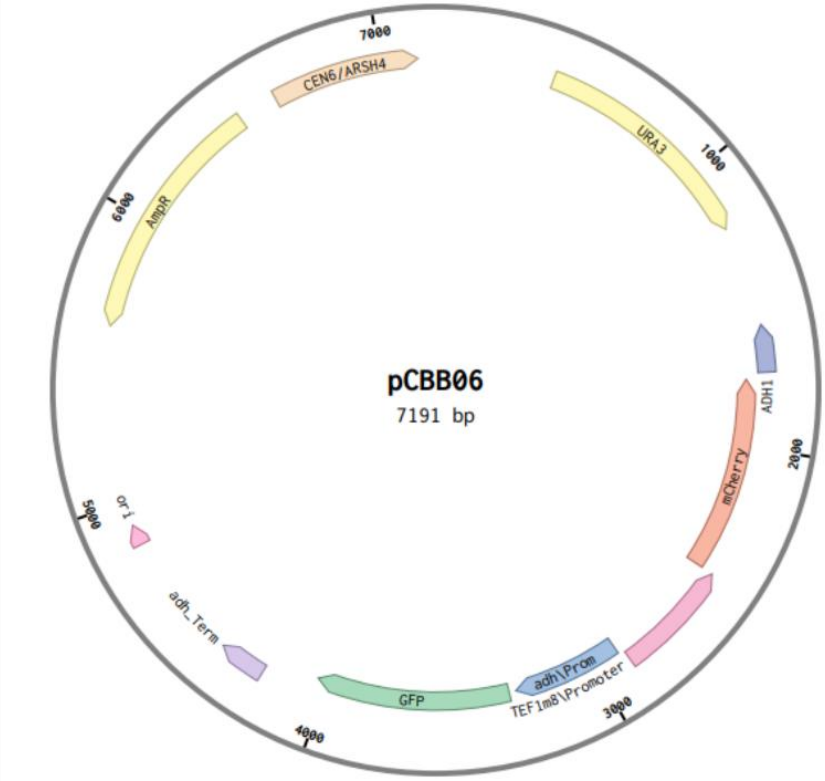

Figure 3.25. Plasmid map of pCBB06. The plasmid map shows schematically the construction of the pCBB06 plasmid used for in vivo screening. The library was inserted before the GFP gene (green) and under the control of the ADH promotor (blue) by the digestion of the plasmid by AgeI (3317) and NheI (3323). GFP expression was normalized by the mCherry expression (orange) controlled by a TEF1m8 promotor. The gene URA3 (yellow) is the selection marker in yeast coding for orotidine 5'-phosphate decarboxylase that allows growth on synthetic media without uracil. The ampicillin resistance gene (AmpR) was used as selection marker for E. coli.

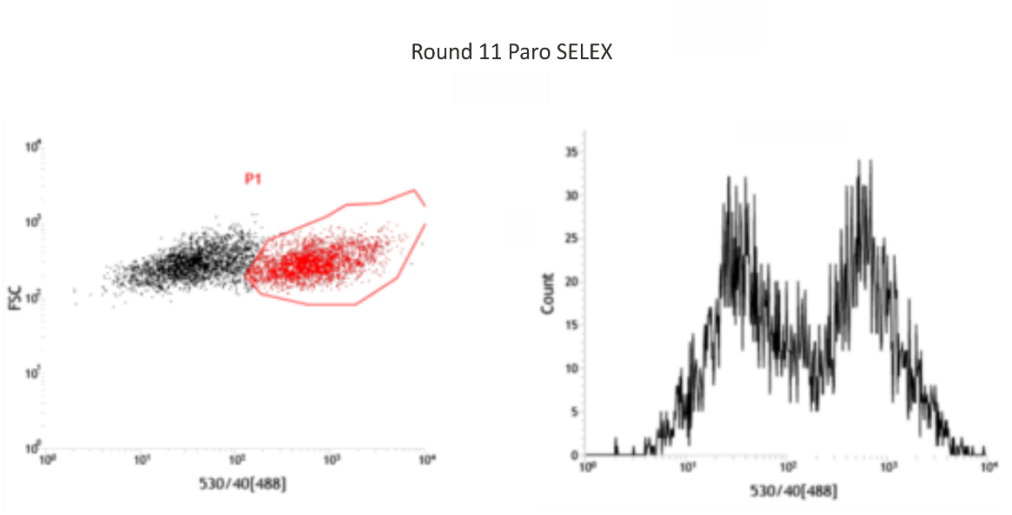

Figure 3.26. FACS sorting of the round 11 cloned library in yeast. In the top graph, the forward scatter is plotted against the fluorescence. Two populations can be clearly observed split approximatively in half each (values in the bottom graph. The dots in black represent the non-fluorescent cells which are discarded. The dots in red represent the fluorescent cells. These cells were recovered and plated on SCD-ura plates to be further screened for riboswitch searching.

For *in vivo* screening, GFP positive sorted cells were plated on SCD-ura plates. After two days of incubation at 30°C, colonies were picked and transfer to a 96 well-plates containing 200 µL of SCD-ura liquid media. After 24 h of incubation at 30°C with a shaking of 1200 rpm, 20 µL of cells from each well were split into a plate containing 180 µL of SCD-ura either without antibiotic or with 100 µM of PARO and incubated for another 24 h in the same conditions. Afterwards, 20 µL of cells were transferred into 180 µL PBS and both GFP and mCherry fluorescence were measured using flow-cytometry. For a total ~1200 clones from both rounds (512 from round 8 and 682 from round 11) fluorescence was measured in the absence and in the presence of PARO and the regulatory activity was calculated (figure 3.27.). The candidate P11.2_H2 (found in the round 11) showed the best regulatory activity with a 1.8-fold decrease in GFP expression and was used for further optimization.

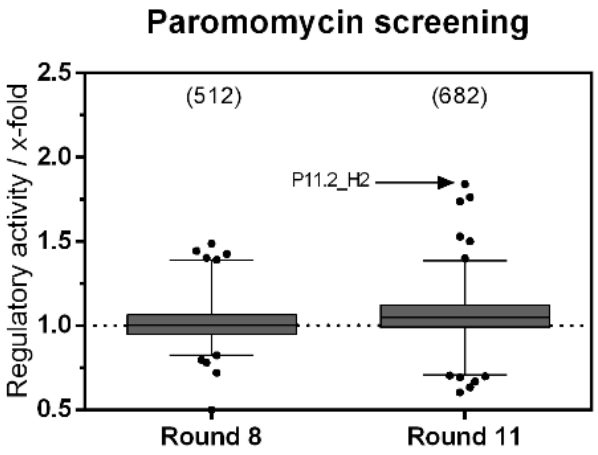

Figure 3.27. In vivo screening for the PARO riboswitch. (A) The boxplot shows the different regulation factors of the GFP gene of the candidates analysed in 96 well plates (number candidates analysed between brackets) in round 8 (left) and in round 11 (right).

3.3.2. Characterization of the P11.2_H2 candidate

The secondary structure of the P11.2_H2 candidates was investigated. The RNA was *in vitro* transcribed, radiolabelled and subjected to an in-line probing analysis. The cleavage pattern is shown in figure 3.28. It suggests that the structure consists of a 7 nt stem with a pentaloop (SL1), a "spacer" region which is most probably unstructured due to its sequence made only of C and A, a second longer asymmetrical stem loop (SL2), with a hexaloop on top, an eleven nt long stem with one bulge on each side and an internal asymmetric bulge with 1 nt on one side and 5 on the other (AB). This loop is closed by a double C-G base pair. A change in the probing pattern shows that this asymmetric bulge may represent the binding pocket (the decrease in flexibility of nt A84, G85, A87 and G88 upon ligand addition) in the in-line reaction. Interestingly, it has been proved very often that the nucleotides responsible for aminoglycoside binding form an asymmetrical internal loop (118). Furthermore, an additional increase in flexibility upon ligand addition is observed in the first stem S1. The nucleotides involved are the three last on the 3' side of the loop (nt 42-44) and the three highest on the 5' side of the loop (nt 30-32). This information suggests that the structure flexibility of this riboswitch may be more complex than expected.

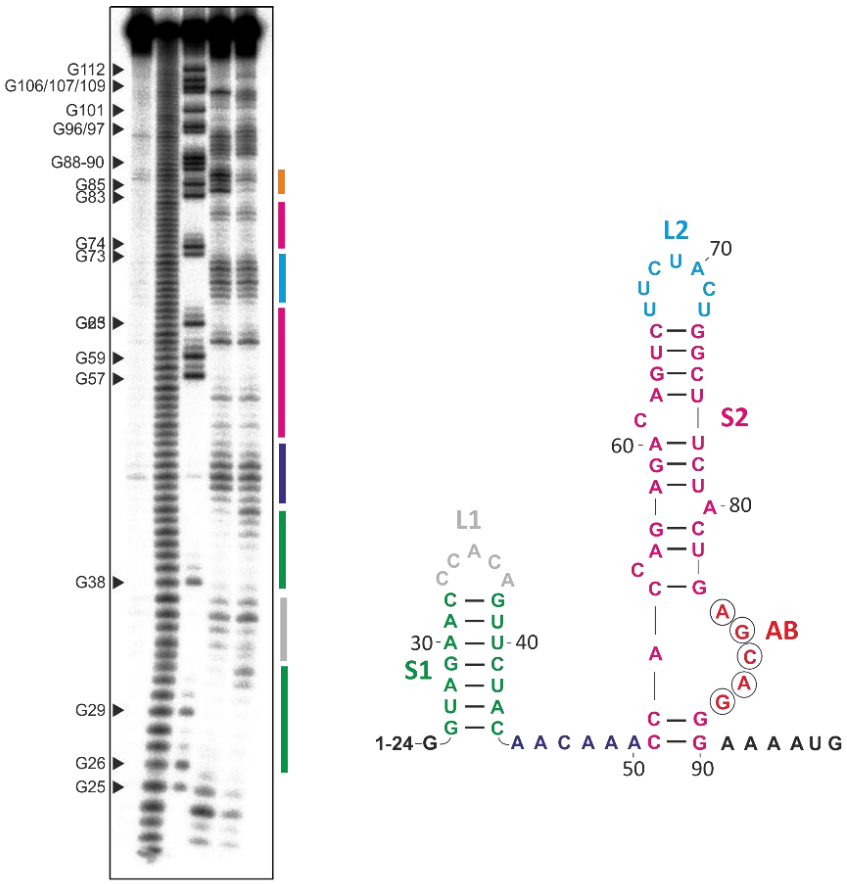

Figure 3.28. Structure determination and characterization of the P11.2_H2 candidate. (A) In-line probing experiment for P11.2_H2 riboswitch. Shown is the cleavage pattern in the absence (–) and presence (+) of 10 µM PARO under alkaline conditions. As references and for nucleotide position assignment, non-reacted RNA (NR), hydroxyl reaction (OH) and nuclease T1 digestion (T1) were loaded onto the gel. G nucleotides and nucleotides that showed a change upon ligand addition are highlighted. Colour coding for identified stem and loop regions follows the coding for the proposed secondary structure in B. (B) Proposed secondary structure of the P11.2_H2 riboswitch including two stems, two loop regions, an asymmetric internal bulge and a spacer region. Nucleotides with changes in the probing pattern are encircled.

To check if the binding site is only in the second stem, the first stem was either removed or shuffled to check if it has any impact on the switching factor. As observed on figure 3.29, removing or shuffling the stem does not remove the riboswitch properties of the P11.2_H2, but it makes the regulation factor slightly worse. For this reason, the stem was not remove for further engineering.

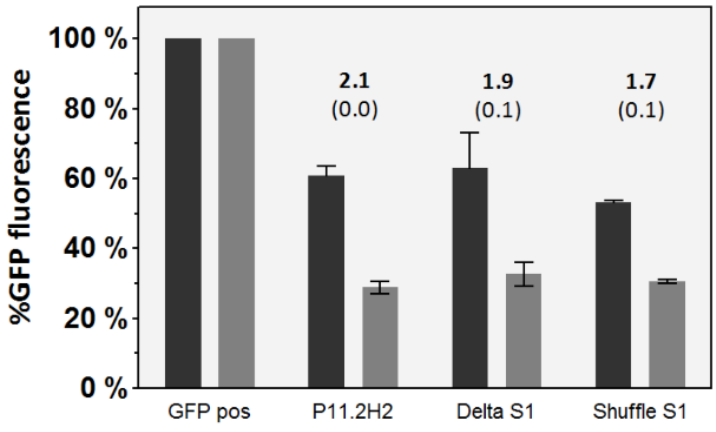

Figure 3.29. Comparison of the GFP gene regulation between the P11.2_H2 candidate and the constructs Delta S1 and Shuffle S1. As observed, none of the constructions delete the switching properties, but slightly reduce the switching factor. This indicates that the S1 plays probably a minor role in the switching properties of the riboswitch.

The dissociation constant of the riboswitch was determined by Isothermal Titration Calorimetry (ITC) (figure 3.30). The analysis resulted in a K_D of 21 nM for PARO. The specificity of the riboswitch towards PARO was tested against NEO. The result shows a K_D for NEO above 10,000 nM, indicating that there is no specific binding. This result explains why only PARO can trigger the riboswitch, only because of the hydroxyl group on the carbon 6 of the first hexose. This indicates that the *in vitro* counter selection is a powerful tool to select highly specific aptamers.

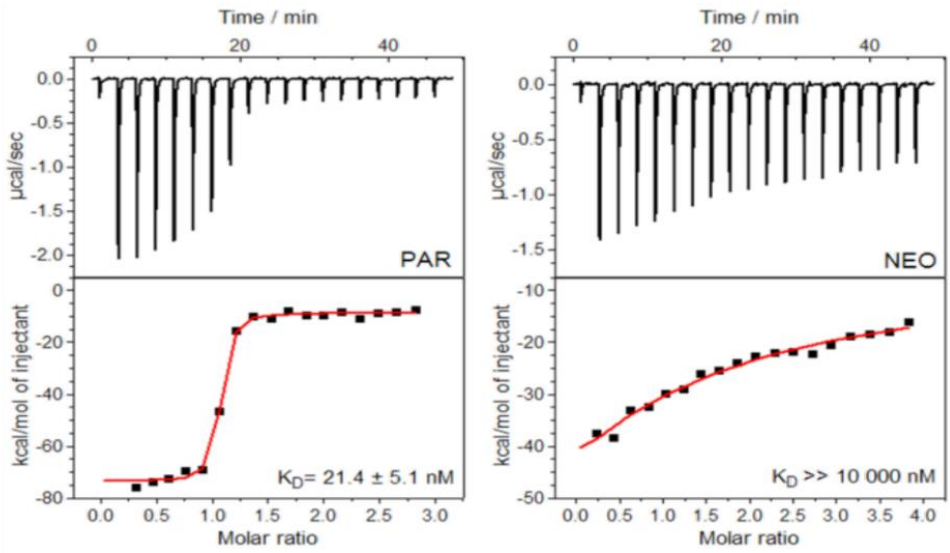

Figure 3.30. Verification of the binding constant of P11.2_H2 with ITC against PARO (left) and NEO (right). Left panel: power required to maintain the temperature of the RNA solution recorded over the time until saturation was reached (baseline-corrected). Right panel: integrated heats of interaction plotted against the molar ratio of ligand over RNA and fitted to a single binding site model (MicroCal PEAQ-ITC Analysis Software 1.1.0).

Other compounds from the same family have been tested against the P11.2_H2 by binding assays similar to the Capture-SELEX protocol (figure 3.31). The results confirmed that the P11.2_H2 candidate only respond to paromomycin. Hence, we decided to continue to work with this candidate.

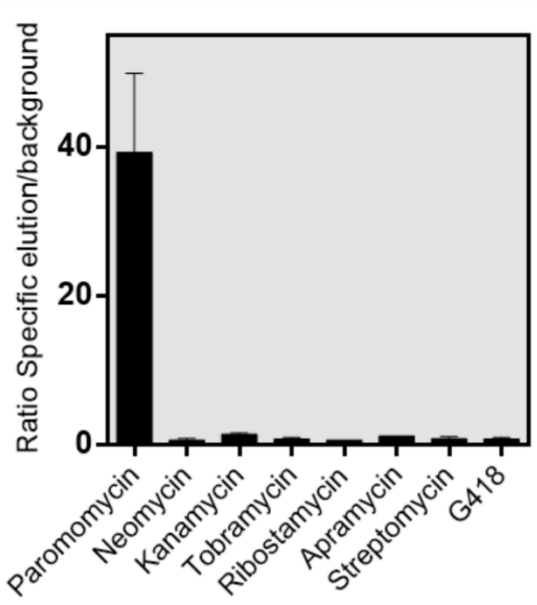

Figure 3.31. *In vitro* specificity binding study of P11.2_H2. The binding of P11.2_H2 candidate was tested by radioactive labelling and immobilization on the beads like in the SELEX protocol. Eight elution were performed in parallel with eight different aminoglycoside antibiotics at 1 mM of target concentration. The ratio between the specific elution and the non-specific elution (measured in the W3 fraction) is showed on the graph. Only PARO is able to undock the RNA from the beads, all the other antibiotics are not able to significantly elute more RNA than only buffer.

To ensure that our candidate is not only active in context of GFP, the P11.2_H2 candidate was cloned in the pCBB05 (pCBB06 + ATG inserted in front of the NheI restriction site, expressing constitutively GFP) in front of the mCherry gene and the fluorescence was compared with and without PARO. The regulation factor observed with 250 µM of PARO is 1.9-fold which is slightly lower than the regulation of the GFP (figure 3.32). The insertion of P11.2_H2 reduced the basal expression of the mCherry to 5% of the control expression level unlike for the GFP, where the basal expression is at 60%. The expression level so close to the background leaving not a lot of room for regulation can explain why the riboswitch is less effective for mCherry as for GFP. But the gene can be regulated. Here we showed that the P11.2_H2 is a PARO specific riboswitch, that can have a regulation effect on two different genes which makes it the perfect candidate for further engineering.

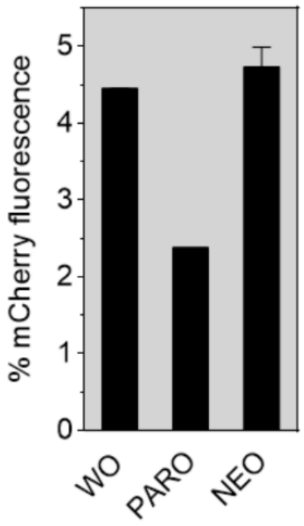

Figure 3.32. Regulation of the mCherry protein expression using the P11.2_H2 riboswitch. The expression of mCherry was compared without ligand (left) with PARO (middle) and with NEO (right). Despite a reduced expression level down to 4.5 %, the riboswitch is still active at a 1.9-fold regulation factor and still specific to PARO.

3.3.3. *In vivo* screening of a doped pool

In order to improve the regulatory activity of the P11.2_H2 a doped screening was performed. The experiment consisted in introducing a few amount of mutations in the P11.2_H2 a sequence the candidates that showed a change in the regulatory activity. Candidate P11.2_H2 was then partially randomized on each position between the first nucleotide in the 5' constant region till the AAAATG (Kozac-start sequence (119)) on the 3' side. Based on our previous work (8), we choose 3.0% randomization of the pool in order to have most of the library containing between one and three point mutations. From this library, 1,532 clones were analysed (figure 3.33) until 100 clones could be identified as gain-of-function (GOF) mutants, meaning a higher regulatory activity than the original P11.2_H2. Additionally, 200 clones were chosen that belong to the class of loss-of-function (LOF), meaning a loss in regulatory activity but still having a comparable GFP expression to P11.2_H2. The sequencing results revealed that three regions contained almost no mutations in the GOF group but concentrate most of the LOF mutants (nt 38-42, 50-63 and 73-90), whereas two regions (nt 33-37 and 64-72) seems to concentrate most of the GOF mutation (figure 3.33).

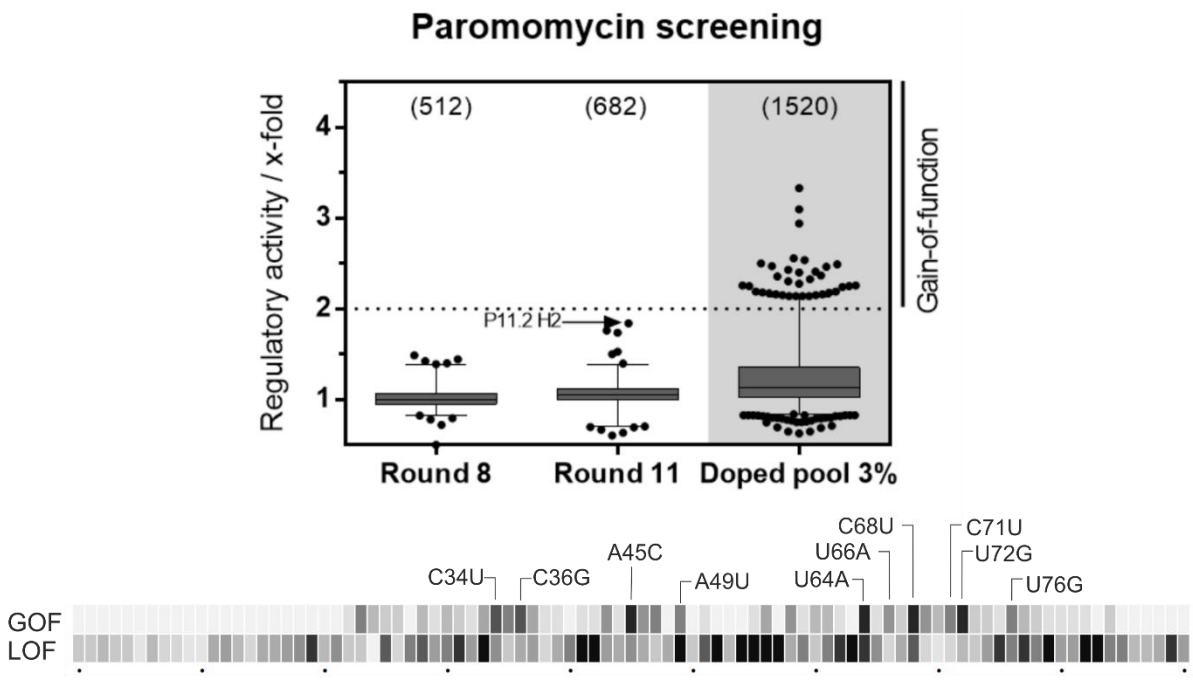

Figure 3.33. The candidate P11.2_H2 showing the best regulation and specificity to PARO was used to generate a doped pool (top). The sequence was synthesized allowing a mutation rate of 3 percent and was screened like previously in vivo (right) in order to find gain of function mutants. All the candidates showing a higher regulatory activity were considered as a gain of function (GOF). (B) The heat map represents the frequency of mutation found in every position depending if they were find in a gain-of-function or a loss-of-function mutant. Most frequent mutations are highlighted.

Furthermore, the ten most frequent mutations found in the GOF were selected (C34U, C36A, A45C, A49U, U64C, U66A, C68U, C71U, U72G and U76G) and were later separately tested (figure 3.34.).

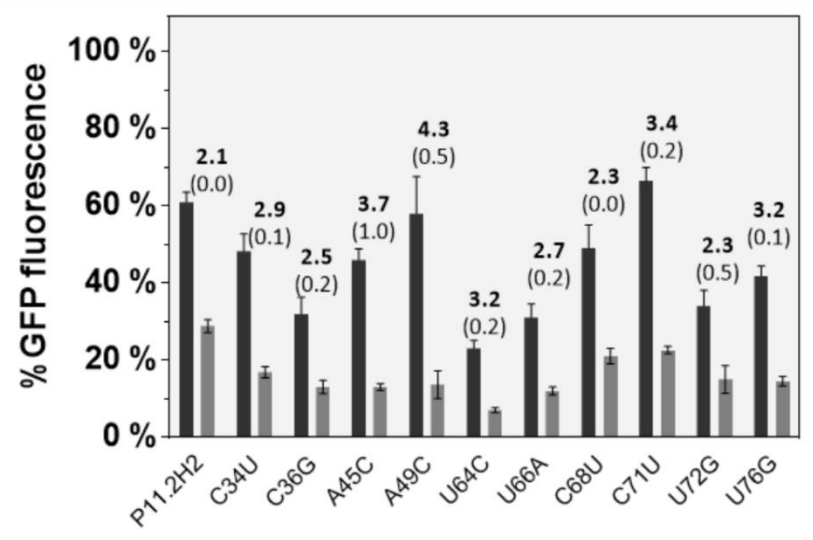

Figure 3.34. The ten most frequent GOF mutations were generated de novo and cloned into yeast. The GFP expression level was compared with PARO (grey bars) and without PARO (black bars). All of the constructs tested showed an improvement in the regulation factor (written above the bars).

3.3.4. Engineering the paromomycin riboswitch

Among all the mutations found in the GOF candidates, the ten most frequent were taken and analysed separately to identify the impact of these mutations independently. Most of the GOF mutations are found because they cause a higher stability in the structure (ex: U64C change a G-U base pair to a more stable G-C base pair). On the other hand, some mutations are both increasing of the on state and a decreasing of the off state, making these mutations more interesting for the design of the riboswitch. None of the mutation is responsible for a significant improvement of the regulatory factor. Therefore, combinations of GOF mutations were designed to engineer the PARO riboswitch. Several combinations were tested (listed in table 4), including sometimes a deletion of the first stem (figure 3.35). The constructs were assembled by amplification of two primers with an overlap of 30 nt and 30-40 nt overhang to allow further homologous recombination on pCBB06.

Table 4: Description of the mutations of the CGOF

Name of the construct	Mutations
CGOF1	A35G, A37G, A49U, U64C, U66A
CGOF2	A49U, U64C
CGOF3	A49U, U64C, U66A
CGOF4	Deletion [28:46], A49U, U72G
CGOF5	Deletion [28:46], U72G
CGOF6	Deletion [28:49]
CGOF7	Deletion [28:46], Deletion [59:78] insertion GAAA between A58 and U79
CGOF8	Deletion S1, A49U, U64C, U66A
PARO-RBS	A45C, A49U, C71U

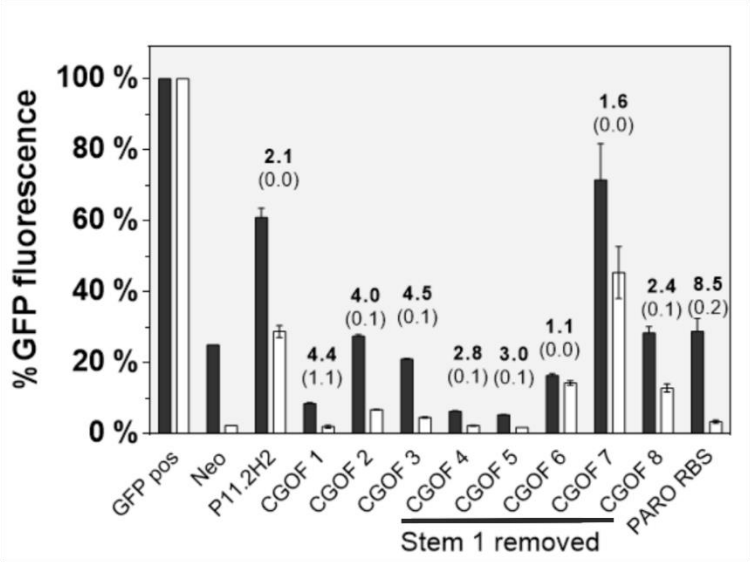

Figure 3.35. Test of several combinations of mutations of the P11.2_H2 candidate. 9 different combinations of mutations were tried and their fluorescence without PARO (black) was compared to the fluorescence with 250 µM of PARO (white). Not all the combinations lead to an improved switching factor. The best combination could achieve an 8.5-fold regulation factor of the GFP gene.

The construct that shows the best regulation factor is the action of the mutations A45C, A49U and C71U (now called PARO RBS) which achieve an 8.5-fold regulation factor with 250 µM PARO and is still not triggered by NEO (figure 3.36). Taken together, initial screening, subsequent partial randomization and the combination of beneficial mutations led to the new synthetic PARO

riboswitch. With 8.5-fold regulatory activity, the dynamic range is comparable to other synthetic riboswitches, e.g. the tetracycline, neomycin or ciprofloxacin riboswitch (8, 90, 91) and other RNA-based devices that control gene expression in eukaryotes.

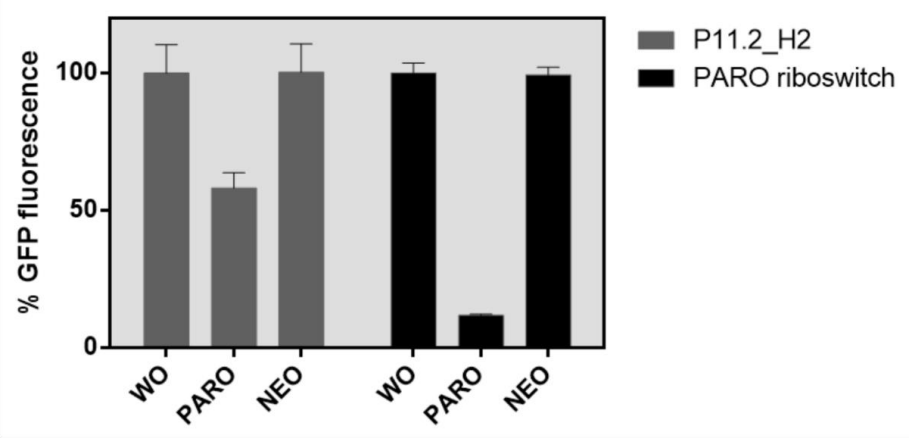

Figure 3.36. Comparison of the PARO-RBS and the P11.2_H2. The GFP expression is compared between the PARO-RBS (black) and the P11.2_H2 candidate (grey) without ligand (left), with 250 µM PARO (centre) and with 250 µM NEO (right). The results were both normalized to 100 % based on their respective GFP expression without ligand. In both case, only PARO could trigger the switch. The switching factor was improved from 2-fold to 8.5-fold with the combination of the three mutations A45C, A49U and C71U.

The dose response activity of the PARO-RBS was tested with three different concentrations of PARO in the media (figure 3.37.). As observed in the figure, the inhibition gets stronger with higher concentration, until 14,3-fold for 1 mM. Furthermore, the PARO-RBS is triggered to 2,7-fold by even only 10 µM of PARO in the media. This indicates an important sensitivity of the riboswitch able to sense lower concentration of antibiotic.

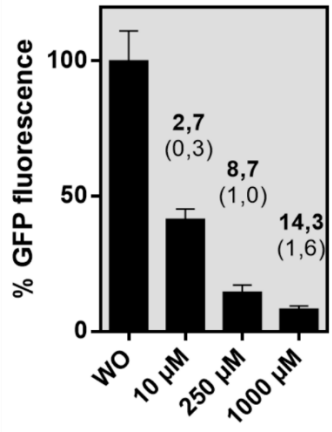

Figure 3.37. Dose dependence experiment of the PARO-RBS. The PARO-RBS regulatory activity was measured compared to different concentration (10 µM, 250 µM and 1000 µM) of PARO in the media.

Finally, the structure of the PARO-RBS was investigated by in-line probing and compared to the P11.2_H2 (figure 3.38.). The in line experiment does not show any major differences between the P11.2_H2 and the PARO-RBS. Interestingly, the flexibility increase of the A43 upon ligand binding is clearly observed. However according to the structure predicted, this nucleotide should not be single stranded. This result suggests the possibility of an alternative refolding of the riboswitch upon ligand binding that create a different structure where this nucleotide is free. This could also explain the increase flexibility observed in the position 30-32 and 42-44.

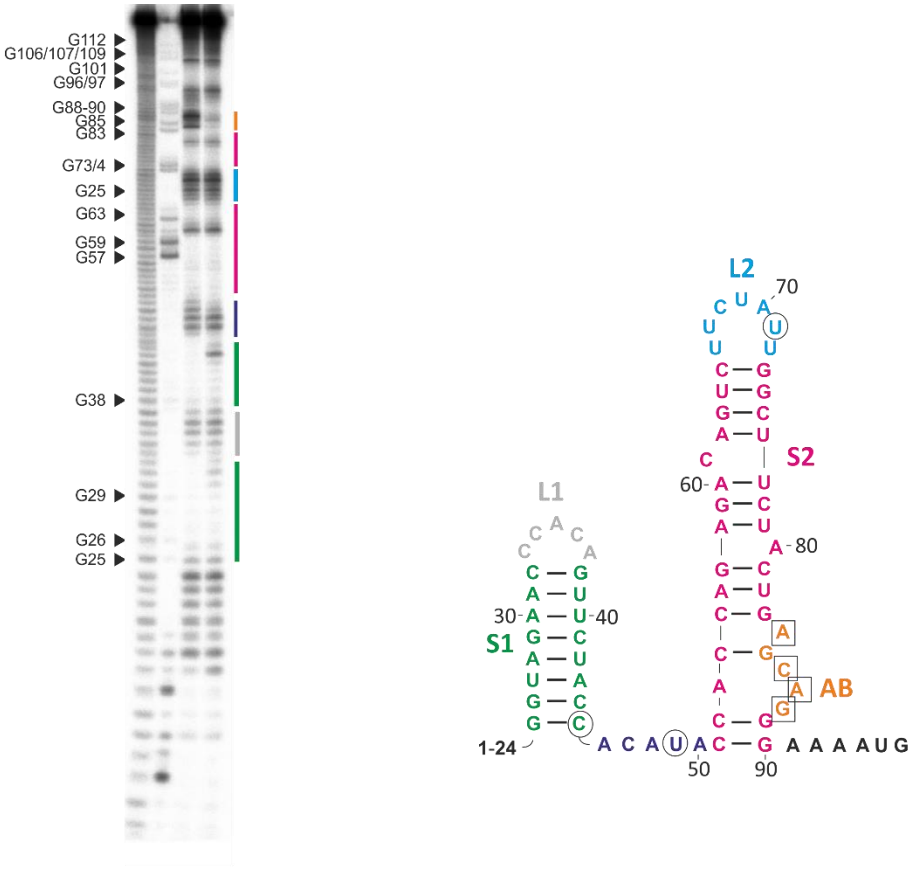

Figure 3.38. Structure determination and characterization of the PARO-RBS. (A) In-line probing experiment for PARO-RBS. Shown is the cleavage pattern in the absence (–) and presence (+) of 10 µM PARO under alkaline conditions. As references and for nucleotide position assignment, non-reacted RNA (NR), hydroxyl reaction (OH) and nuclease T1 digestion (T1) were loaded onto the gel. G nucleotides are highlighted. Colour coding for identified stem and loop regions follows the coding for the proposed secondary structure in B. (B) Proposed secondary structure of the PARO-RBS riboswitch including two stems, two loop regions, an asymmetric internal bulge and a spacer region. The mutated nucleotides compared to P11.2_H2 are encircled.

3.3.5. Target specificity *in vivo* and construction of a NOR repression logic gate

After determining the specificity of the P11.2_H2 candidate *in vitro*, the PARO riboswitch was tested against the same aminoglycosides antibiotics as in the *in vitro* study and its response *in vivo* was analysed to ensure that the introduced mutations have no effect on the specificity (figure 3.39). The light switching observed by geneticin (grey dot) is due to its toxicity to eukaryotic cells. Geneticin has been shown to bind directly the eukaryotic 80S ribosomal complex (120).

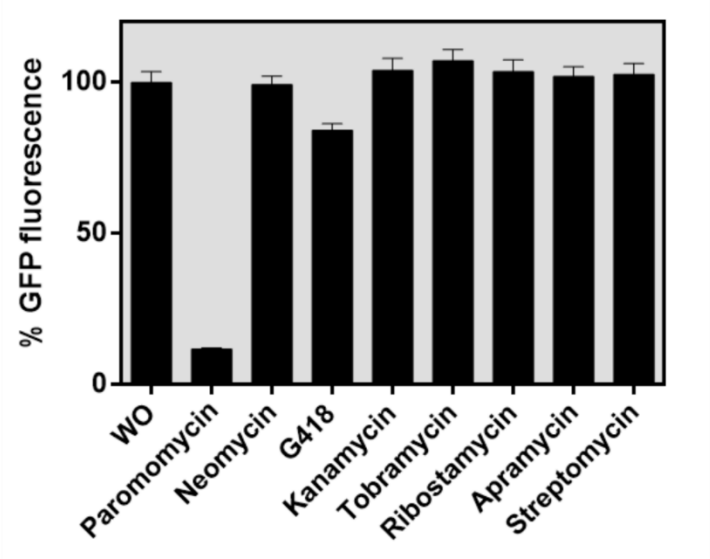

Figure 3.39. In vivo specificity of the riboswitch. The PARO-RBS was tested in yeast against the same range of aminoglycosides as in the binding studies before at a 250 µM concentration. The switching factor is plotted on the arrow. Like in the binding studies, only PARO is able to interact with the riboswitch. The grey dot is the value for gentamicin which is slightly above 1 due to is toxicity for yeast that caused a decrease in GFP expression.

As the graphic shows, only PARO can trigger the riboswitch despite the small differences between these antibiotics indicating the high specificity of our riboswitch. The switch can therefore be use for the establishment of a Boolean gate using the neomycin- (aptamer M4 (91)) binding riboswitch based on the work of Schneider *et al* (7). As we observed that the deletion of the S1 of the PARO RBS was not deleting the switching property, we decided to replace it by the NEO riboswitch. The two stems are separated by the small spacer already present in the PARO riboswitch and the construct was cloned in front of the GFP reporter gene in the pCBB06 (NEO-PARO). Another construct was engineered where the S2 from the PARO-RBS is in front of the NEO-RBS (PARO-NEO). Both riboswitches are split by a repetition of three CAA nucleotides. A last construct was realized where both full riboswitches were placed next to each other. In first position the NEO-RBS, then 7

CAA repeats and finally the full PARO-RBS (NEO-full PARO). All construct are schematically represented in figure 3.40.

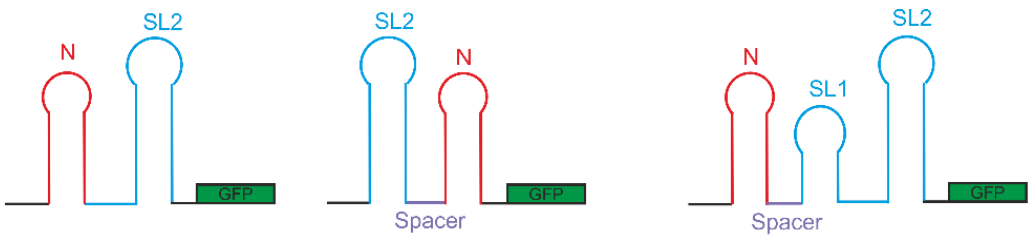

Figure 3.40. Representation of the construct NEO-PARO (left), PARO-NEO (middle) and NEO-full PARO (right). The neomycin riboswitch (N) is represented in red. The paromomycin riboswitch containing the first stem loop (SL1) and second stem loop (SL2) in blue. A CAA spacer region was sometimes introduced and is represented in purple. All these construct were inserted in front of the GFP gene.

Expression level of the GFP of these construct was compared without, with 250 µM NEO, with 250 µM PARO and with both antibiotics at 250 µM (figure 3.41.). The construct NEO-full PARO is too stable and has therefore a very low basal expression making the regulation range small. The NEO-PARO construct behave like expected, showing a regulation for both PARO and NEO when they are added separately (3.5-fold and 5.6-fold respectively) and a higher regulatory activity when both antibiotics are present (15.9-fold).

However, the PARO-NEO shows a very weak regulation with only PARO (1.5-fold) but a very strong regulation with NEO alone (13.2-fold). Furthermore, the regulatory activity with both antibiotics is not so high compared to NEO alone (16,3-fold) and in the same range of the NEO-PARO construct. The NEO-PARO construct shows the expected behaviour for a NOR logic gate and will be therefore use for further experiments.

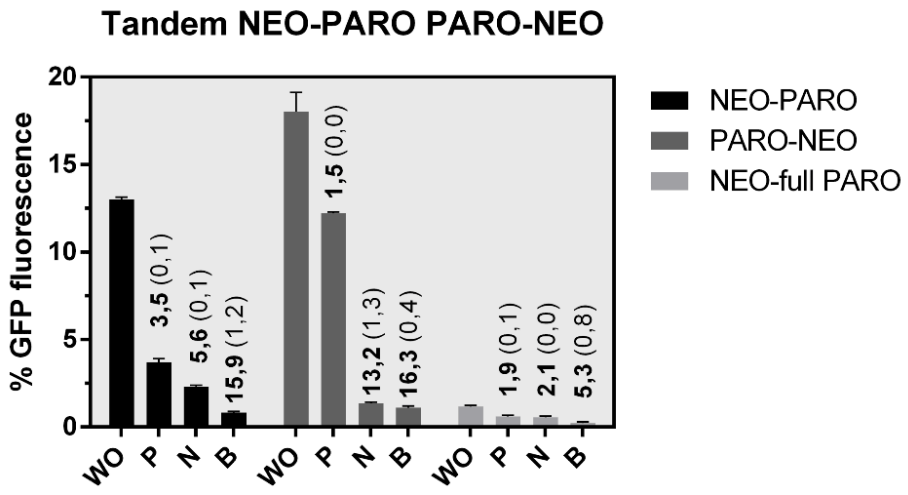

Figure 3.41. Comparison of the GFP expression of three tandem constructs NEO-PARO (black), PARO-NEO (dark grey) and NEO-full PARO (light grey): Experiment was made either without ligand (WO), with 250 µM PARO (P), with 250 µM NEO (N) or with 250 µM of BOTH antibiotic (B). The switching factor of each construct depending of the antibiotic in the media is written above the bar with standard deviation between brackets.

The basal expression of the NEO-PARO construct is reduced to 13% due to the insertion of heavy structured RNA sequences imposed by the riboswitches, but still leaving room for testing an ON/OFF inverter gate. On a first level, the device can thus be used to only moderately inhibit gene expression, whereas its application as a NOR gate leads to a multiplied decrease of residual reporter expression to about 0,7% with both PARO and NEO, yielding an increased level of regulation. To characterize the behaviour of this construct depending of the concentration of both antibiotics, the switching factor was calculated depending of the concentration of each target in the media (figure 3.42.). A good correlation is observed between the increasing of the ligand concentration and the switching factor. A switching factor of 32-fold could be achieved with both target input at 1 mM. Interestingly, both target at small concentration have a higher regulatory activity (20 µM of each make a 4.85-fold switching) than one of the target alone (4.5-fold for 1000 µM of PARO alone and 5-fold for 500 µM of NEO alone).

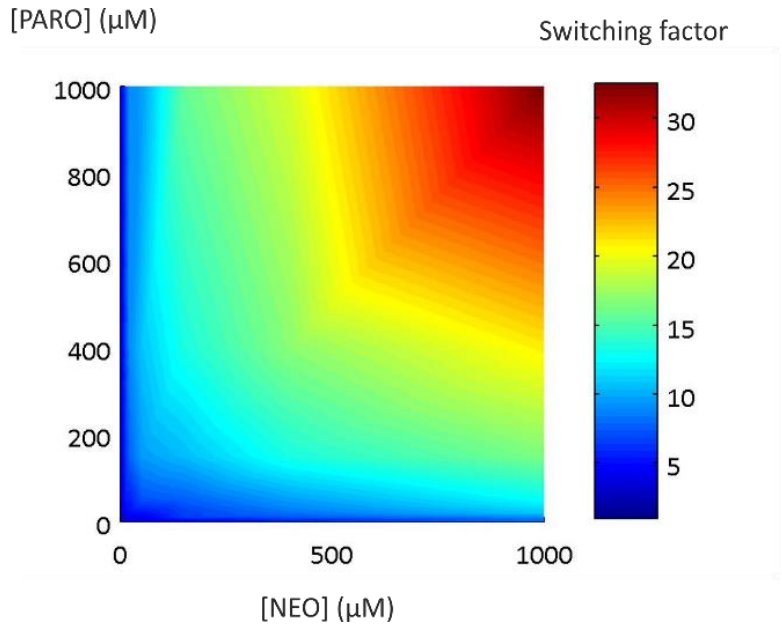

Figure 3.42. Dose response experiment of the NEO-PARO construct in function of the PARO and NEO concentration. The darker blue the colour is, the lower the switching factor is. The redder, the higher is the regulatory activity.

The NEO-PARO construct was also tested against different combinations of antibiotic at 250 µM for each target and the switching factor was plotted on a heat map (figure 3.43). The NEO-PARO construct shows expected behaviour depending of the ligand added to the media. As the NEO switch is as well able to be slightly triggered by RIBO, the NEO-PARO construct is also slimly affected by the presence of RIBO. The combination of both NEO and PARO in the media causes the best switching factor up to 13.8-fold.

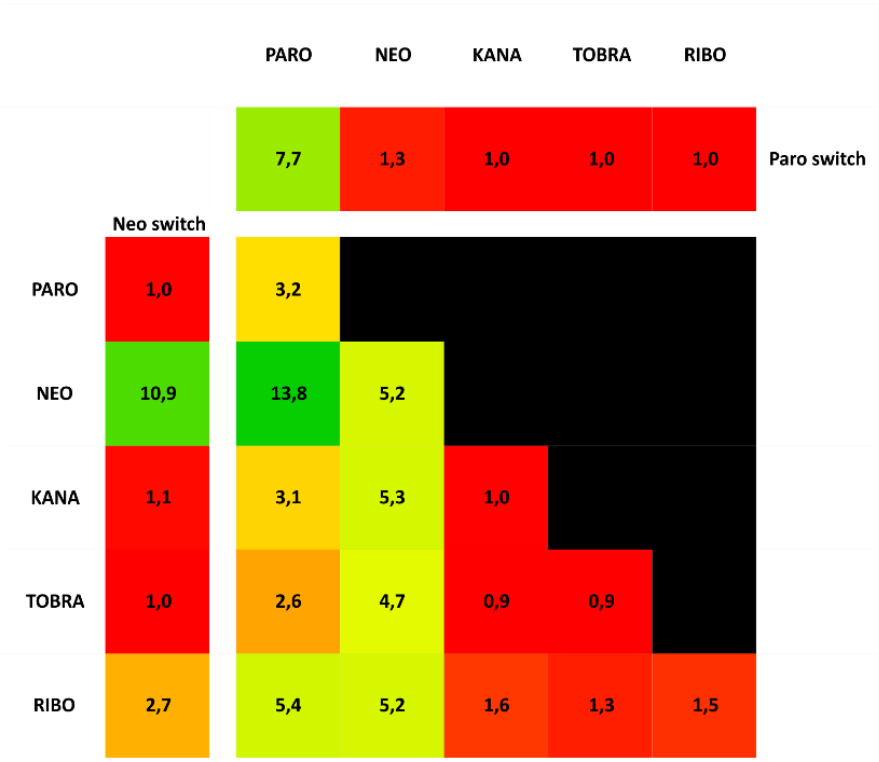

Figure 3.43. Heat map of the switching factor of the PARO-RBS (first line), the NEO-RBS (first column) and the logic gate NEO-PARO. Both riboswitches were tested against 250 µM of each of these antibiotics. For the PARO switch, only PARO triggered the switch. However, for NEO, RIBO is also able to slightly trigger the switch. The logic gate NEO-PARO was tested against each five antibiotics previously tested, but as well on all the combination of two different antibiotics at 250 µM each. The lower the switching factor is, the more the color of the square will turn red. The higher the switching factor is, the more it will tend to green. As expected, the NEO-PARO shows the best switching factor when the combination of both NEO and PARO are added to the media.

As we already prove that only the asymmetric bulge is sufficient to have a switching activity, a construct was engineered where there are two times the binding pocket incorporated (one on top of the other). By including two binding sites, our hypothesis was to decrease the off-state without decreasing too much the on-state. The design of the construct and its expression level of GFP without and with PARO is represented on figure 3.44.

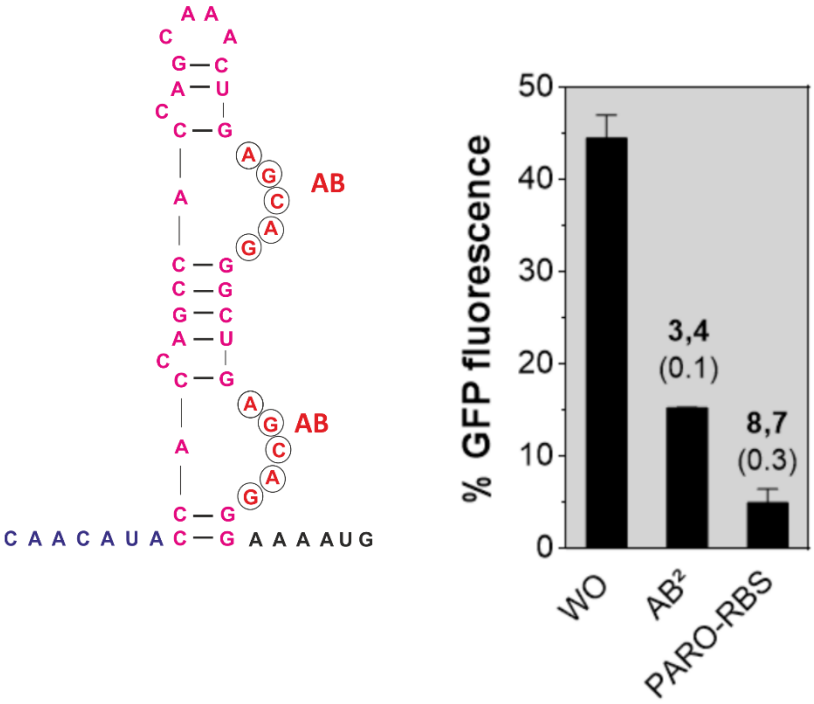

Figure 3.44. **Test of the double asymmetric bulge construct.** Structure of the double asymmetric bulge construct (left). The construct was tested with 250 µM PARO and the expression without ligand (WO) and with PARO (PARO) was compared. The switching factor is written in bold and the standard deviation between brackets.

4. Discussion

4.1. Evaluation of the development of the Capture-SELEX for RNA

The development of the Capture-SELEX protocol adapted for RNA was not only a simple change from DNA to RNA from the original protocol (71) but the adaptation of every step to optimize it for our purposes: the development of riboswitches. The first step was a design of a library that has to be suitable for *in vivo* screening. The library was designed first to regulate splicing upon ligand binding by the inclusion of splice sites inside the library, based on an already existing splicing controlling riboswitch (92, 121). Indeed, based on the folding prediction of the aptamers found in the DNA Capture-SELEX (115), these aptamers use their DS within a stem loop structure for refolding to get removed from the CO when bound to the ligand. Riboswitches that regulate splicing in yeast showed the same characteristic. Therefore, the library was designed as a full intron, where the extremities of the random regions were designed based on the 5' splice site on the 5' end and on the branch point/3' splice site for the DS and the 3' end. The choose of the size of the random region was also optimized based on an optimal length to find aptamers (122) but also to fit to a realistic size of an intron. This design showed that it could be spliced out in the absence of the ligand when inserted inside a GFP gene.

The CO design showed that this is a critical step. To allow the hybridization of the CO on the pool plenty of parameters have to be taking into account (123). Magnesium and RNA concentrations appeared to be parameters that have an influence on the immobilization yield. However, the extension of only one G of the CO showed a stronger binding to the library despite a very low increase of the melting temperature. This result suggests that with the addition of this nucleotide, we are very close to a dissociation of the RNA from the CO but still more RNA is bound to the CO, suggesting that even aptamers with a high K_D could get dissociated from the DS to bind their target. Indeed, the melting temperature of the CO has to be certainly high to really capture RNA, but on the other hand, if the sequence is too long and the melting temperature too high, the affinity to the CO of the library will be too important to allow dissociation of the complex to bind target with a K_D in the micromolar range, which is often the case for small molecules (62). Using a biotin-streptavidin interaction for the immobilization of the complex CO-RNA on the beads is the most suitable solution for this SELEX. The binding happens without any chemical input that remove the risk of any interference caused by the addition of chemical reagent or the use of non-physiological conditions.

The folding protocol was also adapted for RNA. RNA is much more subject to degradation than DNA especially at high concentrations of magnesium due to its reactive 2'OH group on the ribose. Therefore, RNA can't be boiled at high temperature with alkali or heavy metal ions (124). The

protocol was adjusted to be performed with magnesium and CO together, to include the presence of the CO as a folding possibility for RNA but without hydrolysing the RNA. A temperature of 65°C for 5 min is a good compromise to unfold RNA without degrading it and a slow cool down to room temperature (21°C) on the thermoblock allows RNA to explore all possible conformation without being trapped in a local minimum of energy where it could may be not interact with the CO.

Using NEO as proof of concept target lead to the first enrichment using our Capture-SELEX protocol. The choice of the target was critical specially to test a complete new protocol. NEO was choose because of several reason that makes it a good target. For example, NEO possesses numerous amino groups, it binds to the bacterial ribosome and also aptamers were already selected against it with a very fast enrichment (72). The Capture-SELEX performed showed an enrichment after 6 rounds of selection indicating already a good protocol able to enrich aptamers in a few rounds. The specificity is also present in the pool, as KANA is barely able to elute RNA and other compounds except NEO are totally unable to elute RNA specifically. The protocol was also successfully applied to a SELEX with mixture of ligand containing PARO, AMOX and CIT but only PARO could be enriched from the library. This result confirmed that some targets are more predominant than others to become enriched during the SELEX. Indeed, the fact that only one target enriched instead of a little bit of each proves that the speed of enrichment is different for each target, depending of the "average K_D" of the pool to the target (125). The average K_D means the mean value of the distribution of all the K_D values of all RNAs present in the naïve pool. If a starting pool contains 100 binders with an affinity to target A of 1 nM but only 1 binder with an affinity to target B of 1 nM because of their difference into their "average K_D" to the library, the probability of enriching the target B without losing it during the selection process is much lower than to enrich the library for target A.

In vivo screening showed that no riboswitch that could regulate splicing could be identified. The idea was probably too optimistic as splicing is a complex process where several enzymes are involved. For the regulation of splicing by the tetracycline aptamer, the position of the splice site next to the riboswitch was very critical. Indeed, a shift of one or two nucleotides killed the regulatory activity. The probability that a regulating aptamer was selected that in addition exactly captured the splicing site enough during the refolding is too low to be probably achieved in this way. However, screening for translation regulation in yeast does not involve the sequestration of a specific sequence but only the formation of a more stable structure that can make the ribosome fall off the mRNA (126). Therefore, we switched for this strategy.

4.2. Evaluation of the optimization of the RNA Capture-SELEX

The RNA protocol established, and the new pool design was a success. They allowed the enrichment of three RNA libraries against two aminoglycoside antibiotics (PARO and KANA) and a peptidic antibiotic (VIO). The difference between the amount of RNA undocked at the elution step compared to the amount of RNA eluted at the last wash step demonstrated that the protocol generated RNA sequences that have enough affinities to get removed from the DS and stay bound to the free target. Furthermore, all these aptamers must be "structure-switching" aptamer due to this undocking from the CO. Therefore, these libraries are suitable for *in vivo* screening to find active riboswitches that interfere with translation initiation.

However, the protocol showed also that not any target could enrich a pool enough to have an observable difference using a radioactive tracer. Both penicillin and metabolites as targets showed no difference in the specific elution and in the background elution. More surprisingly, the same result was observed for ATP and CMP. Despite their propensity to interact with RNA in their natural function (107), no enrichment could be observed using Capture-SELEX. When a target is not able to enrich a pool, there are some explanation.

The first one is, all the potential good binders were lost during the selection. Indeed, the efficiency of immobilization on the beads and on the CO is below 100% (based on the data from the first round, only 20% of the RNA are still on the beads after all the washing steps), which means that some sequences are lost during this step. Additionally, if we take the case of one good binder among the pool, there is a competition between the CO and the target for binding. As this RNA can't be bound on both the target and the CO, there is a probability that this binder doesn't undock from the CO and get lost in the selection process (125). Based on the binding studies of the candidate P11.2_H2, that has a K_D of 21 nM for PARO, only 60% of the RNA were undocked from the beads after a 5 min elution despite the excess of target of around 100.000-fold. This result shows that even good binders can be lost because of probabilities that could not happen. This probability is even lower when some targets can't generate aptamers with a sufficiently good K_D. Therefore, the protocol needs some optimization to increase this probability.

The second one is due to the too weak partitioning effect. This phenomenon could be observed when the washing step was still 30 min. Indeed, the longer the washing is, the more it causes unspecific undocking of a lot of RNA from the CO, suggesting that most of RNA eluted during the selection was only unspecific undocking. During the first round of SELEX we have a lot of RNA still immobilized on the beads that could resist to all the wash steps. If we consider that only 10^{14} different sequences are still on the beads at the elution step (~1/10 of the starting library), with a Ks

of 10^{-12} M (affinity of the binding between an RNA sequence and the CO), we have theoretically around $3,6*10^7$ sequences eluted non-specifically just to reach the equilibrium in the ideal case. However, we can estimate the real quantity of RNA eluted by RT-PCR. By experience, we know that a concentration of 20 pmol/µL of RNA makes a clear and strong band on the gel after the RT-PCR step. In the first round of Capture-SELEX, a band of the same intensity was always observed only 4 cycles and a reaction divided in three tubes. This indicates a concentration of about 240 pmol/µL eluted from the beads equivalent at $7*10^{12}$ different sequences. If we consider the amount of real binders inside of the library at least a 1000 time less frequent than the theoretical non-specific undocking, we have a partitioning effect 200.000 worse than the theoretical one. In the following rounds, not more than 6 cycles of RT-PCR were done, indicating as well a high RNA concentration in the elution fraction. With a such high amount of RNA, the number of PCR cycle to perform after the RT must be kept low to avoid overamplification phenomenon, which prevent the good binders to have chance to really get amplified and to see a clear enrichment in the further rounds.

Despite its easy and straightforward protocol, the Capture-SELEX was not working in all cases and needed optimization. So far, three of them were investigated in this thesis. All of them was targeting the major drawback of this protocol, the low partitioning. Indeed, unlike in the classical SELEX where the RNA bind the ligand on the target a first time and then bind again the free ligand added in excess in solution to the matrix (8, 62, 72), Capture-SELEX only has one selection step (undocking from the CO and binding to the target). This has to be compensate by a strict stringency to lower to the minimum the amount of non-wanted RNA recovery. As observed before, the stringency applied was too weak to allow enrichment for some target to be visible despite the presence of real aptamer in the pool that could be detected after a higher stringency.

However, the lack of information about the content of the pool makes it harder to tune the stringency of the SELEX accurately. There is unfortunately no way to identify during the selection step when we remove enough non-wanted RNA as much as there is no way to determine when the SELEX is too stringent. Concerning the time of elution, the amount of non-wanted RNA washed away increases linearly with time whereas the aptamers are supposed to bind in the very first seconds. Indeed, based on ITC measurements (127), the binding reaction happens in seconds and the equilibrium is reached again. According to these affirmations, a better partitioning could be achieved by only resuspending the beads for a few seconds with your target in solution, and directly recover the SE fraction afterwards. Theoretically, very few of non-wanted RNA will be undock while all the binders should bind to the target. Moreover, this strategy could aim for aptamer with a very good

k_{on} allowing solely the fast binders to get selected. On the other hand, the k_{off} will not be submitted to selection. Therefore, RNA with an average or bad K_D could be enriched.

For the last improvements done in this protocol, an enrichment could be finally visible for harder targets. Indeed, GLY is a very small molecule, that reduces the amount of possible interaction with RNA compared to aminoglycoside antibiotics. FBP was also tried several times with the old protocol without any success. This last results shows that again the improvement added to the protocol helps to get an enrichment of a library in the SELEX procedure. The weak partitioning was also confirmed as the main issue to get an enrichment. Despite the heating step, the longer CO and the very short elution time the partitioning is not perfect. The pool is probably containing a half of non-binding RNA that survive throughout the SELEX. Reducing the non-specific elution to zero is impossible because of the binding equilibrium between the CO and the RNA pool. However, any further improvement to this protocol has to be make in this direction. First of all, a better understanding of all the steps of the Capture-SELEX have to be studied in details and optimized. For example, which CO sequence achieve the best ratio between low non-specific elution and undocking of aptamers when the target is present, how many washes, at which temperature and for how long are necessary to remove the biggest part of background RNA, etc… Another improvement which has to be realized is about the stringency for better K_D. Some of the aptamers found with the Capture-SELEX enriched almost all the library but got a K_D in the micromolar range (71) which is for some application unusable. Indeed, most of all the *in vivo* active synthetic riboswitch have a K_D below the micromolar range until the low nanomolar range. However, the enriched pool from PARO was able to deliver an *in vivo* active aptamer with a low nanomolar range K_D.

4.3. Regulation of gene activity by a paromomycin riboswitch

A functional and specific riboswitch towards PARO could be engineered starting from an enriched library from the Capture-SELEX. Several results observed here proved that the Capture-SELEX protocol is more adapted than the classical SELEX approach to develop synthetic riboswitches. First of all, the hit rate of the *in vivo* screening to find riboswitches was higher than in the classical SELEX. We can use for comparison the data from the ciprofloxacin riboswitch paper (8). In this paper, 5000 candidates were screened and only one was able to switch. In this work, only 1200 were screened and 10 candidates were showing riboswitching properties.

Next generation sequencing data revealed that our candidates P11.2_H2 was found as the 65th most enriched sequence in the round 11. Further investigations showed also that the most enriched sequence (which represent 40% of all the library) of this round is the same sequence with a unique

point mutation on the last random nucleotide and 15% of the top 100 sequence are derived from this sequence. All this information brought together proves that the Capture-SELEX protocol is a major improvement for the discovery of new functional riboswitch. Indeed, not only it can select and enrich aptamer easier, faster for target that are not suitable for standard SELEX, but also can specifically enrich for aptamer with riboswitching properties which need structural change to be functional (128).

A comparison can be made with the *in vivo* screening performed for the neomycin riboswitch. In our case, half of the sequences inserted in the 5'UTR of the GFP allowed fluorescence detection. For the neomycin screening, only 5% of the candidates allowed gene expression (91). In the case of the neomycin, the SELEX was performed in the classical way, with immobilization of the target on a matrix and elution of the binders by addition in excess of the target. Despite the generation of good binding aptamers, this technique was generating mostly too stable RNA structure, annihilating any gene expression. Furthermore, the global expression level of our best candidates is relatively high (more than 25% of the constitutively expressed GFP control, 60% for the P11.2_H2) compared to the 8.2% of the N1 candidate from the neomycin screening (91) or the 4.5% of the 10A candidate from the ciprofloxacin screening (8). This data shows the Capture-SELEX strategy is selecting and amplifying aptamers with riboswitch properties, by forcing the aptamer to be unfold without ligand (binding to the CO means that the docking sequence as to be free) but completely refolding in the presence of the ligand (undocking of the sequence meaning the formation of stable structure where the DS is involved). Preliminary data showed already that a TOBRA riboswitch could be found directly from the *in vivo* screening with a switching factor of around 5-fold and a K_D of 1 nM with several binding sites showing again the superiority of Capture-SELEX to develop synthetic riboswitches. Indeed, by forcing to enrich aptamers that have high flexibility (unfolded without target, folded with target), aptamers selected with Capture-SELEX possess the same characteristic of the natural and synthetic riboswitches, refolding upon ligand binding (93, 129)

In addition, despite its high basal GFP expression, the P11.2_H2 has a K_D in the low nanomolar range, belonging to the same range of affinity as the neomycin (91), tetracycline (90) and ciprofloxacin riboswitch (8). This results stresses that despite the lack of a two-step selection. Capture-SELEX possesses only one step where the real selection happens, when the target is put in solution with the library and binders have to compete between CO and target. This makes the theoretical stringency lower than for the classical SELEX where the aptamer has to bind a first time to the target on the matrix, and then bound to the free target in solution. One binding selection step lower the stringency for Capture-SELEX, but aptamers generated by Capture-SELEX can have a very high affinity

towards their targets in the same range as other aptamers and riboswitches (130). The specificity of the P11.2_H2 is even higher than the NEO or CIPRO riboswitch for example. Indeed, the NEO switch has a switching factor of almost 3-fold with 250 µM of ribostamycin. On the other hand, PARO only shows a binding *in vitro* of only PARO but also *in vivo* by switching only with PARO with a high switching factor. These results demonstrate that there is a direct link between specificity and affinity (131). Indeed, when high specificity is acquired by the use of counter-elution (61, 62), it means that all reactive group that can interact with the aptamer is part of the binding. It has been proved that changing moieties on a target of an aptamer decreases its affinity to this target (8). Reaching a high specificity for aptamers, which is something that can be perform easily with the Capture-SELEX due to the binding of the target happens in solution, can drive the library towards the enrichment of high affinity and high specificity aptamers.

The generation of a doped pool to increase the riboswitch efficiency has been already used to find the CIPRO riboswitch and was applied for the P11.2_H2 candidate. This low rate of mutation allows the change of one to three nucleotides in the riboswitch and identification of better candidate can be done using the same *in vivo* screening protocol (8). The screening of 1520 candidates revealed 100 mutants with gain-of-function and 200 candidates with loss-of-function. All these candidates were sequenced and directly highlighted area that could be modified for the improvement of the riboswitch. Indeed, by analysing the frequency of these mutations, hotspots can be identified. This could confirm that any mutation close to the binding pocket leads to a complete loss of riboswitching properties, like observed for the adenine riboswitch (132). The PARO riboswitch was assembled by bringing together mutations that cause the best improvement of the switch but also the ones that do not decrease too much the ON-state. Indeed, a too low ON-state by a combination of stabilizing mutations will most often lead to a too stable switch where the ribosome could not read through at any state. The A45C mutation is causing the formation of a longer stem in S1 allowing the base pairing with the G25 leading to a slightly lower ON-state but also a lower OFF-state. The stem 1 has been proved to be helping for the switching and apparently when more stable the switching gets improved. This stem probably is here to slow down the ribosome scanning and if the ribosome become slower next to the binding pocket, once the PARO binds to the S2 and stabilize it more, the ribosome fall off. However, in absence of PARO, the second stem is more flexible and allows the ribosome to read through. The A49U mutation appeared in the spacer region between the two stems. Our hypothesis is that the region from 47-49 become a CAU sequence that can base pair with the AUG on the other side of S2 and are brought together only when PARO stabilize the lower part of the S2. Hence, only the OFF-state is affected by this mutation. The sequestration of the start codon

could be also one of the reason of the effectiveness of this mutation. Indeed, hiding crucial sequence for translation or for splicing is the mechanism of a lot of riboswitches (133). The C71U mutation is however not perfectly understood. Indeed, this mutation appears in the L2 far from the binding site and does not alter the structure of the switch according to mfold predictions. Though, the mutation is in the docking sequence which is supposed to be partially complementary of the 5' constant region. By mutating one of the nucleotide, this structure is may be not stable enough to form *in vivo* and was may be having a bad influence on the riboswitch.

The combination of all these mutations led to the creation of an 8.7-fold riboswitch with 250 µM of PARO. The optimization of the riboswitch allowed even to see a clear switching behaviour upon addition of only 10 µM, making this switch able to sense small amount of metabolite like natural riboswitches (134). The mutations introduced has no effect on the affinity of the riboswitch to PARO because they are far away from the binding site and do not change the K_D. However, these mutations bring the structure of the riboswitch to a state where only a slight input brought by the PARO is able to transform the structure into a stable state that is upon target binding too strong for the ribosome. That is the key of a good functioning riboswitch. There is a correlation between the delta G (or stability) of the structure put in the 5'UTR and the expression level. However, this correlation is not linear but more look like a sigmoid where only next to certain value a change in expression level can happen. If the delta G is too high, small variations (like the binding of the target on the riboswitch) will not cause a change big enough to cause a visible effect because the ribosome could still scan through this structure. The same case applies for too stable structure, if the delta G is too low allowing only barely gene expression, the input of the target binding cannot bring much lower the gene expression as it was already really weak before. However, when the delta G of the riboswitch is just next to the limit of what the ribosome can scan through, the input of the target that binds to the riboswitch and stabilize it will bring the delta G low enough to be too stable to allow a good scanning off the ribosome (135).

The proof that S1 does not prevent the switching but only lower it has been made in this thesis. Therefore, this stem could be in theory changed by another one and the switching properties should stay the same. If the S1 is changed by another riboswitch, it could work as a standard Boolean NOR gate. Such construct has been already shown a tight gene regulation depending on the target input (7). Hence, S1 was changed by the NEO riboswitch. The switching factor observed in all the cases (NEO, PARO or both) behave like expected. Indeed, the NEO riboswitch is a bit more efficient than the PARO riboswitch and in this construct as well. But their values are lower than when the switch are alone. When both input are given, the expression level drops below 1%, meaning almost a

complete turn off of the gene. Furthermore, this construct has been proven to be the most active in the presence of both NEO and PARO. A slight drawback is that the NEO riboswitch is not specific only to NEO but also is triggered slightly by RIBO (136). RIBO is slightly able to compensate NEO but very weakly in this construct. The dose response of this gate shows a perfectly increase of the switching factor when the concentration of the inputs is higher. Interestingly, the input of two antibiotics even at a lower concentration is more efficient than a high input of only one antibiotic. For example, 50 µM of each antibiotic causes a switching factor of 7,2-fold which is 1.6-times higher than 1 mM of PARO alone and only slightly lower than 1 mM of only NEO. It seems that to achieve a better switching factor, both riboswitch needs to be folded around their targets. Indeed, if one of the switch is fully on the ON-state, this can be easier for the ribosome to scan instead of having only one switch folded. The assembly of the tandem NEO-PARO shows that gene expression can be precisely modulated with the input of one, the other or both antibiotics allowing downstream application for sensitive gene expression tuning but also complete shutdown of the gene controlled by the construct. The PARO-NEO construct however shows a complete different behaviour. Indeed, both input at 250 µM shows a regulatory activity in the same range as NEO-PARO. Though, PARO is barely able to regulate the gene expression alone but on the other hand NEO is almost as efficient as both antibiotics together. The results suggest that when the S2 is moved away from the ATG, it loses a lot of this regulating ability. Assembly of two different riboswitches in a row is not always a plug-and-play system. High RNA structures are inserted in the 5' UTR and complex folding can happen leading to a complete misfolding of the original riboswitch causing a loss of activity for one or both of them (7). However, the presence of the S2 seems to slow down enough the ribosomal scanning enough without reducing to much the basal expression to allow a better regulation of the NEO riboswitch alone. A try was made to include the full PARO riboswitch after the NEO riboswitch. Theoretically, it should have increase the regulatory activity of the PARO switch alone now that all of the riboswitch is present. Though, the introduction of the full construct lead to a very low basal expression without ligand, showing a very poor regulatory activity in all the cases. This result demonstrates the importance of finding the "sweet-spot" for gene regulation (137), where the insertion of a sequence in the 5' UTR is not too stable to allow gene expression, but close enough of the strength where the ribosome can't read through anymore (8).

The inline probing experiment and the CGOF8 construct showed that the AB plus the spacer region are responsible for the binding and the switching of the ligand. Our hypothesis was that, if another binding site was added to the AB, it could improve the switching factor by creating a bigger gap between the delta G without and with PARO. The result obtained confirmed that theory. Indeed,

putting a second AB on the first one had for consequence to more than double the switching factor. Given that the ON-state of this construct is at 45%, it could be used in a combination of a third binding site. However, cloning repeats can be challenging due to unspecific binding of primers during the PCR step. The construct is though fitting perfectly to the construction of another logic gate by being engineered with another riboswitch.

4.4. Evaluation of a C-less SELEX

Interestingly, the C-less SELEX against NEO was unsuccessful despite the choice of a good target. One explanation could be the modification of the docking sequence lead to a poor yield of immobilization despite that the calculated melting temperature did not changed. The amount of RNA loaded on the beads on every round was in the same range as all the other SELEX performed in this thesis, excluding the theory of the low immobilization rate.

The most probable theory is that despite the increase of flexibility in the library due to a very low amount of C in the sequence (only two in the DS and some in the primer binding site) the RNA is not able to unfold from the CO. Indeed, the energy brought by the GC base pairing is strong and can help to refold the RNA around the target. With only the possibility of forming weak stems, the RNA is not anymore able to undock from the CO where GC base pair are present. The observed results confirmed this theory. Indeed, the ration SE/W3 used in every SELEX to check the presence of an enrichment (which is supposed to be around 1 without enrichment) was always below 0.5. This indicates that a very few amount of RNA is eluted. In addition, the number of PCR cycles performed after the SE to observed a visible band on the control gel (normally always between 4-14 cycles) was always more than 16 even in the latest round, indicating a very poor amount of RNA undock during the elution and the absence of the enrichment.

C-less SELEX may be an interesting subject to investigate more in detail due to the characteristic and advantage it can offer. However, this technique does not seem suitable with our current Capture-SELEX protocol.

5. Methods

5.1. General protocols for Capture-SELEX

5.1.1. Taq-PCR for pool preparation

A large volume PCR was used to amplify the template oligonucleotide by the addition of both forward and reverse primer to produce enough copies of each single template present in the template oligonucleotide into 4 different 96 well plates compatible with the PCR equipment (total volume 41.5 mL). The general protocol is described here below.

Table 5.1. PCR mix

Reagent	Concentration	Volume (µL)	Final concentration
Taq-Buffer	10 x	4150	1 x
Forward primer	100 µM	830	2 µM
Reverse primer	100 µM	830	2 µM
Template	100 µM	33,2	80 nM
dNTPs	100 mM each	62,25	1,5 mM
Taq polymerase	100 x	415	1 x
MQ-H_20	100%	34,85	-

Table 5.2. Thermocycler program

Step	Duration (s)	Temperature	
Initial denaturation	180	95°C	
Denaturation	60	95°C	} 5 cycles
Annealing	60	58°C	
Elongation	120	72°C	
Cool-down	Infinite	4°C	

5.1.2. *In vitro* transcription

For the pool preparation, after purification of the PCR template by sodium acetate precipitation, DNA was solubilized in 2000 µL of water. From this template, one tenth was used as a template for the T7 polymerase to transcribe RNA. Based on the number of cycles, this amount is enough to cover the diversity needed for the SELEX. The reaction was incubated overnight at 37°C. The reaction mix is described in table 3.

Table 5.3. Transcription mix for RNA library production

Reagent	Concentration	Volume (µL)	Final concentration
$Mg(Ac)_2$	1 M	200	20 mM
Tris HCl pH=8	1 M	2000	200 mM
DTT	1 M	40	400 µM
Spermidin	200 mM	100	2 mM
NTPs	100 mM each	400	4 mM
T7 RNA polymerase	1000x	10	1 x
RNAse Inhibitor (moloX)	2500x	4	1x
DNA template	90 µM	200	1,8 µM
MQ-H_2O	100%	5772	-
		Σ=10.000	

After the transcription, pyrophosphate residues were pelleted by a 1 min centrifugation at 9000 rpm and the supernatant was transferred into a 50 mL screw-cap tube. Reaction was afterwards stopped by the addition of 2 mL of EDTA (0,5 M pH = 8), 1 mL of NaAc (3M pH = 6,5) and filled up to 40 mL with a 1:1 mixture of ethanol/acetone. The RNA was stored 1 h at -20°C and then was centrifuged for 90 min at 10500 rpm. The supernatant was afterwards discarded and the pellet washed two times with 10 mL of 70% ethanol for 15 min. After the last wash, the maximum possible of supernatant is remove with the pipette and the RNA is afterwards dried 10 min at room temperature. RNA was solved in 4 mL of water and 4 mL of 2x RNA loading dye.

For the transcription step in the SELEX, a different mix is used and is described in table 4. This mix is used to generate RNA from the RT-PCR step for the next SELEX round. The RT-PCR product is directly used as a template without any purification or quantification as long as it produces a strong single band on the agarose gel performed after the RT-PCR.

Table 5.4. mix for transcription used in the SELEX

Reagent	Concentration	Volume (µL)	Final concentration
Tris HCl pH=8	1 M	4	40 mM
DTT	1 M	0,5	5 mM
NTPs	100 mM	2,5 each	2,5 mM
$MgCl_2$	1 M	1,5	15 mM
MQ-H_2O	100%	71	-
RT-PCR product	-	10	-
^{32}P-α-UTP	3,3 µM	1	33 nM
T7 RNA polymerase	50 U/µL	2	100 U
RNAse Inhibitor (moloX)	4000 U/µL	1	40 U

5.1.3. DNA/RNA purification – Sodium acetate/Ammonium acetate/butanol

Sodium acetate precipitation: 1/10 vol. 3 M sodium acetate and 2,5 vol. ethanol 100% were added to the DNA/RNA solution and mixed. The mixture is incubated at -20°C for 1h, centrifuged at 4°C at 13000 rpm for 30 min: The supernatant is discarded and 800 µL of 70% ethanol is added and vortex with the pellet followed by a 15 min centrifugation step at 13000 rpm. RNA is dried 10 min at RT to evaporate the rest of ethanol. The pellet is resolved in H_2O and stored at -20°C.

Ammonium acetate precipitation: ½ vol. 7,5 M ammonium acetate and 2,5 vol. of ethanol 100% were added to the RNA solution and mixed. The mixture is incubated 10 min on ice and centrifuged at 13000 rpm at room temperature for 15 min. The supernatant is discarded and 200 µL of 70 ethanol is used to wash the pellet followed by a 5 min centrifugation step at 13000 rpm at room temperature. The wash step is repeated two times and RNA is afterwards dried at 37°C for 5 min. RNA is suspended in 50 µL water and stored at -20°C.

Butan-1-ol precipitation: 10 vol. of butanol is added and mixed strongly for one minute. The sample is directly transferred to centrifugation for 20 min at 13000 rpm at room temperature. Two wash steps were performed before dissolving the pellet in water.

5.1.4. Polyacrylamide gel purification

For RNA purification, a 6% denaturing polyacrylamide gel purification is performed. For each 20x20 cm gel a 100 mL of mixture is prepared. The mixture is detailed in table 5.5.

Table 5.5. mix for polyacrylamide denaturing gel

Reagent	Concentration	Amount	Final concentration
Polyacrylamide	40%	15 mL	6%
Urea	-	42 g	7 M
TBE	10x	5 mL	0,5x
MQ-H_20	100%	Complete till 100 mL	-

Once the gel glass plates are assembled, 800 µL of ammonium persulfate 10% and 80 µL of *N*,*N*,*N*',*N*'-Tetramethylethylenediamine are added to the mixture to start the gel polymerisation. The mixture is poured between the glass plates, then the comb is inserted and gel is let 30 min at room temperature for polymerisation. After removing the comb, the gel is placed on its device and loaded with 2 mL of RNA per gel. The migration condition is set on 20 W per gel and they are run until the lower band of the loading dye reach the buffer chamber. The RNA band is located afterwards by UV shadowing and cut out of the gel. The band is chopped and transfer to a 50 mL screw cap tube containing 20 mL of 300 mM sodium acetate for an overnight incubation at 4°C under mild agitation. Afterwards, the RNA that migrated in the liquid is filtered to remove small pieces of gel and is again precipitated by the addition of 100% ethanol. After two wash steps using 10 mL of 70% ethanol, RNA is solved in 200 µL of MQ-H_20.

5.1.5. Native polyacrylamide gel for EMSA

The elecromobility shift assays (EMSA) are performed using 15% polyacrylamide native gel. The mixture is detailed in table 5.6.

Table 5.6. Mixture for native PAGE

Reagent	Concentration	Amount	Final concentration
Polyacrylamide	40%	22.5 mL	15%
Glycerol	100%	6 mL	10%
TB	10x	6 mL	0.5x
MQ-H_20	100%	25.5 mL	-

5.1.6. Radioactive labelling – Body labelling/End labelling

For the purpose of tracking RNA in each fraction of the SELEX RNA needs to be radioactively labelled. For this purpose, a 10 pmol of the starting RNA library is dephosphorylated using the Antarctic phosphatase, precipitated using butanol precipitation and solved in 5 µL MQ. RNA is phosphorylated again with ^{32}P-γ-ATP using the T4 phosphonucleotide kinase (PNK). RNA is afterwards purified using ammonium acetate precipitation. Protocol for both steps are respectively described in table 5.7 and 5.8.

Table 5.7. Mixture for dephosphorylation

Reagent	Concentration	Volume (µL)	Final concentration
RNA	7,5 pmol/µL	1	0,5 pmol/µL
Phosphatase buffer (Roche)	10x	2	1x
Antartic Phosphatase	1 UI/µL	2	2 UI
MQ-H_20	100%	15	-

Table 5.8. Mixture for phosphorylation

Reagent	Concentration	Volume (µL)	Final concentration
Dephosphorylated RNA	3 pmol/µL	5	0,5 pmol/µL
PNK buffer (Roche)	10x	2	1x
Phosphatase	1 UI/µL	2	2 UI
MQ-H_20	100%	15	-

5.1.7. RT-PCR

After RNA elution in the SELEX, the RNA needs to be transformed into cDNA and to be amplify by PCR. The RNA is first mix with the buffer, then heat up to 65°C for 5 min and afterwards both enzymes are added in the mix. RT-PCR mix and protocol are detailed in table 5.9 and 5.10.

Table 5.9. mixture for RT-PCR

Reagent	Concentration	Volume (µL)	Final concentration
Taq-Buffer	10x	10	1x
First strand buffer (Invitrogen)	5x	4	0,2x
DTT	1 M	0,2	2 mM
Forward primer	100 µM	1	1 µM
Reverse primer	100 µM	1	1 µM
$MgCl_2$	50 mM	3	1,5 mM
dNTP mix	25 mM each	3	0,75 mM
MQ-H_20	100%	25,8	-

Table 5.10. RT-PCR program

Step	Duration (s)	Temperature	
Reverse transcription	600	54°C	} 4-16 cycles
Denaturation	60	95°C	
Annealing	60	58°C	
Elongation	120	72°C	
Cool-down	Infinite	4°C	

5.1.8. Agarose Gel electrophoresis

For DNA extraction and analysis high quality agarose was used. After casting the gels with agarose concentrations either of 1% (for plasmids) or 3% (for PCR products and SELEX) samples were mixed with DNA loading dye pipetted into pre-formed wells alongside of the adequate molecular length marker. Gels were run in 1x TAE buffer at 6V/cm for 20-60 min. Next, gels were stained with 0.5 µg/mL ethidium bromide in water for 5 min. DNA was visualized with 254 nm light for analytical purposes and with 365 nm for extraction.

5.1.9. Scintillation measurement

For the estimation of the amount of RNA present in each fraction, detection and measurement of the radioactivity were performed. 2% of each fraction were transferred into scintillation Pony vials (perkinelmer) containing 2 mL of scintillation cocktail. These vials were transferred into a rack and were read by the machine.

5.2. General protocols for molecular cloning

5.2.1. PCR with Q5 polymerase

PCR with Q5 polymerase was used to amplify a defined sequence from two overlapping primers or from a vector to create a double strand DNA that can be used for homologous recombination in yeast, digestion or sequencing. This polymerase was used because of its proof-reading ability that decrease the number of mistakes during elongation step. Temperature for primer hybridization and durations for primer extension were adjusted to target length and primer-vector hybridization temperatures. Mixture details and protocol are given in table 5.11 and 5.12.

Table 5.11. Q5 PCR mixture

Reagent	Concentration	Volume (µL)	Final concentration
Q5 Buffer	5 x	20	1 x
Forward primer	100 µM	2	2 µM
Reverse primer	100 µM	2	2 µM
GC enhancer	5 x	20	1 x
Template	40 ng/µL	0.5	2 ng/µL
dNTPs	10 mM	2	200 µM
Q5 polymerase	2 U/µL	0.5	1 U
MQ-H_20	100%	53	-

Table **5.12.** Thermocycler program

Step	Duration (s)	Temperature	
Initial denaturation	120	98°C	
Denaturation	10	98°C	
Annealing	60	58-72°C	} 5 cycles
Elongation	2 s/100 nt	72°C	
Final extension	180	72°C	
Cool-down	Infinite	4°C	

5.2.2. Restriction digest

DNA sequences (plasmids or inserts) that has to undergo a ligation were incubated at 37°C between 1h to overnight with two type IIP restriction enzymes. PCR products were purified by column purification. Digested vectors were purified by gel extraction. Restriction enzymes were inactivated by heating step 80°C for 20 min or discarded during purification steps. Digestion was performed in 50 µL. The mixture is detailed in table 5.13.

Table 5.13. Reaction mix digestion

Reagent	Concentration	Volume (µL)	Final concentration
Cutsmart Buffer	10 x	5	1 x
Vector or	-	2-10 µg	
Insert	-	Purified PCR reaction	
Restriction enzyme	20 U/µL	1-4	0.4-1.6 U/µL
Template	40 ng/µL	0.5	2 ng/µL
MQ-H_20	100%	Fill up to 50 µL	-

5.2.3. Ligation

Assembly of a digested insert with a digested backbone was performed either at room temperature for 1h or overnight at 16°C to increase the efficiency and yield. The molar ratio between vector/ insert was chosen between 1:3 and 1:5 depending of the length of the insert. 50 ng was always used for the ligation. The reaction was afterwards purified by butanol precipitation and dissolved in 5 µL water. 4 µL were used for transformation of competent *E.coli* cells. The mixture is detailed in table 5.14.

Table 14: Reaction mix for ligation

Reagent	Concentration	Volume (µL)	Final concentration
T4-DNA ligation buffer	10 x	2	1 x
Linear plasmid	-	50 ng	2.5 ng/µL
Insert	-	Depending on length	-
T4-DNA ligase	400 U/µL	1	20 U/µL
MQ-H_20	100%	Fill up to 20 µL	-

5.2.4. General protocol for yeast homologous recombination

To test our constructs and using the highly efficient DNA repair machinery in yeast, homologous recombination was performed using competent yeast cells. A purified digested backbone and insert which possess a sequence homology of 30-40 nt at the junction of both part. This length is optimal to have high yield in the recombination. The molar ratio vector : insert was set to 1:5 – 1:10. The transformation in competent cells was performed with the Frozen Yeast EZ II Transformation Kit from Zymo Research was chosen to realize all yeast transformations according to the manufacturer's instructions.

5.2.5. Transformation of chemo-competent *E. coli*

Cells were thawed on ice for 15 min. 100 µL of cells was added to the purified ligation reaction and incubated for 1 h on ice. Afterwards, cells were incubated in a 42°C water bath for 60 secs for the heat-choc and put back on ice for 20 min. 1 mL of SOC medium was added to the cells and the tube was transferred into a thermoblock for 1h at 37°C at 1000 rpm agitation. Then, cells were pelleted by a 1 min centrifugation step at 13000 rpm and resolved in 100 µL of SOC medium. Cells were plated on LB plates containing 100 mg/L ampicillin and incubated overnight at 37°C.

5.2.6. DNA isolation – Mini preparation from *E. coli* cells

All plasmids were isolated with Qiagen Plasmid Miniprep Kit. The DNA was eluted in 50 µL MQ-H_20 and stored at -20°C for further use.

5.2.7. DNA isolation – Mini preparation from yeast

To recover plasmid in yeast cells, a 4 mL yeast culture was grown until saturation and cells were harvested by centrifugation 1 min at 13000 rpm. Cells were resuspended in 250 µL of P1 buffer from the Qiagen Plasmid Miniprep Kit and 100 µL of glass beads (diameter 0.25-0.5 mm) were added and vortexed for 5 min. After letting the beads and cell debris settle to the bottom of tube, the supernatant was transferred in another reaction tube and 250 µL of P2 buffer was added following by a mixing with inversion. Afterwards 350 µL of N3 buffer was added and the rest of the protocol

was done following the Qiagen Plasmid Miniprep Kit. DNA was eluted in 20 µL and 5 µL were used for transformation in *E. coli* competent cells.

5.2.8. Nucleic acid quantification

To measure the DNA or RNA concentration, a Nanodrop ND-1000 was used by loading 1.5 µL of nucleic acid on the pedestal that will measure the absorbance at 260 nm. The concentration was determined using the Beer-Lambert's law.

5.2.9. Yeast colony PCR

To amplify the content of the sequence inserted in the 5'UTR of our plasmid, colony PCR was performed on yeast. One colony was transferred into 10 µL of NaOH 20 mM and was boiled for 10 min at 99°C. Afterwards, 1 µL was transferred to a Q5 reaction mix and was amplified for 35 cycles.

5.2.10. Sequence analysis

Once DNA was extracted from cells (plasmid or PCR product) samples were diluted in water with 3 µL of a 10 µM sequencing primer. The samples were next analysed by Seqlab laboratories, Goettingen.

5.2.11. Preparation of chemo-competent *E.coli*

Top 10 strain was streaked out on a LB plate and incubated overnight at 37°C. One colony was afterwards transferred into a 4 mL LB medium tube and incubated at 37°C overnight at 150 rpm. 300 mL of pre-warmed LB medium were inoculated to an OD_{600} of 0.1 and grown until an OD_{600} of 0.6 – 0.8. Cells were cooled down to 4°C on ice and harvested by centrifugation with 4000 rpm for 15 min at 4°C. Cells were washed with 150 mL ice-cold $CaCl_2$ 0.1 M and resuspended in 25 mL 0.1 M $CaCl_2$. Sterile glycerol was used to adjust the cells suspension to 15% glycerol. 330 µL aliquots were flash-frozen in liquid nitrogen and stored at -80°C.

5.2.12. Preparation of chemo-competent yeast

The Frozen Yeast EZ transformation Kit from Zymo Research was chosen to prepare competent yeast cells. The procedure was performed according to manufacturer's instructions.

5.2.13. Yeast glycerol stocks

4 mL of SCD-ura medium were inoculated with a single yeast colony and cultured overnight. 700 µL of this culture was transferred to a sterile 1.5 mL tube and 300 µL of sterile glycerol 50% was added to a final concentration of 15% glycerol. The tube was afterwards stored at -80°C.

5.3. Adapted protocols for molecular cloning and reporter gene analysis

5.3.1. RNA-library cloning for *in vivo* screening

For the *in vivo* screening of a library to find active riboswitches, an optimized protocol has to be used to ensure the highest yield possible of sequence integrated in the plasmid (see figure 5.1. for the details of the location of the insertion). Here below is detailed the homologous recombination protocol for big libraries.

RS453α cells were streaked out from a frozen stock on a YPD plate and incubated at 30°C for 48h. One colony was used to inoculate a 4 mL YPD liquid culture overnight and was used to inoculate a 100 mL YPD liquid culture. The culture was grown until OD_{600} reached 1.6. Cells were centrifuged 3 min at 3000 rpm at 4°C and the supernatant was discarded. Cells were washed two times with 25 mL of ice cold sterile water and centrifuge 3 min at 3000 rpm at 4°C. After discarding the last water fraction, cells were washed one time with electroporation buffer (1 M sorbitol, 1 mM $CaCl_2$) and centrifuged 3 min 3000 rpm at 4°C. The supernatant was discarded and the cells were resuspended in a final volume of 20 mL of LiAc 0.1 M and DTT 10 mM solution. Cells were incubated 30 min at 30°C at 150 rpm in a 100 mL sterile flask. Afterwards cells were spin down 3 min at 3000 rpm at 4°C and washed with 50 mL of electroporation buffer 3 min at 3000 rpm at 4°C, supernatant was discarded. The competent cells were resuspended in 400 µL of electroporation buffer to reach a final volume of 1 mL.

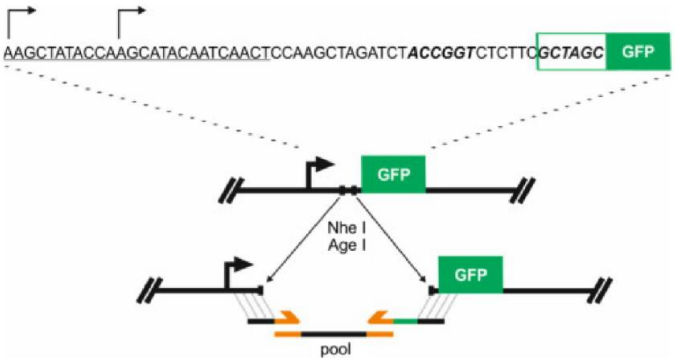

Figure 45.1. Homologous recombination strategy for sequence insertion in the 5'UTR of GFP. The restriction site used for the linearization of the pCBB06 are represented in bold and italicized. Sequences that needs to be inserted are amplified with primers containing between 30-40 overhang nucleotide to the vector upstream of AgeI and downstream NheI. The reverse primer introduces also the kozak-start sequence.

For the homologous recombination 1 µg of vector digested and purified and 2 µg of insert were mixed with 400 µL competent cells. For the controls, 300 ng of vector was used with 100 µL of cells.

Electroporation was performed with a pre-cooled sterile electroporation cuvette and electro-shocked at 2.5 KV and 25 µF for 4 ms. Next cells were carefully transferred into 8 mL of a 1:1 mixture of YPD/Sorbitol 1M (2 mL for controls) and incubated 90 min at 150 rpm. Cells were afterwards spin 3 min at 3000 rpm at room temperature, the supernatant was discarded and cells were resuspended in 1 mL SCD-ura. Cells were split in half and each 500 µL were transferred into 125 mL SCD-ura and incubated 48 h at 150 rpm at 30°C. 1 µL of the cells were used to estimate the transformation efficiency by plating a 1/1000, 1/10000 and 1/100000 dilution on SCD-ura plates. After the 48 h incubation, the two flasks were mixed and 5 mL were transferred into a 45 mL SCD-ura media and incubated 24 h at 150 rpm at 30°C (1/10 dilution). Afterwards, 1.66 mL of the previous culture were transferred into fresh 50 mL SCD-ura and incubated 24 h at 150 rpm at 30°C. Finally, 500 µL were taken out and transferred 50 mL fresh SCD-ura and incubated 24 h at 150 rpm at 30°C.

5.3.2. Fluorescence Activated Cell Sorting for selection of GFP positive cells

Once the cells are transformed with the library, 4 mL of the 50 mL culture is taken out, centrifuged 5 min at 3000 rpm at 4°C. Cells are washed with 4 mL 1x PBS and centrifuged again in the same conditions. Cells are resuspended in 4 mL PBS and filtrated through a 40 µM cell strainer. Cells are kept on ice before FACS. After sorting, cells were plated to achieve a concentration of 200-300 colonies per plate.

5.3.3. *In vivo* screening

Looking for active riboswitch, *in vivo* screening has to be performed. For this, cells were individually picked and transferred into 96 well plates containing 200 µL of SCD-ura. Plates were incubated at 30°C at 1200 rpm for 24 h. Afterwards, 20 µL of this pre-culture was transferred to 180 µL of SCD-ura either without ligand or with 100 µM of ligand and incubated in the sane conditions as the pre-culture. Then, 20 µL of cells were transferred to 180 µL of PBS 1x in a 96 well plate compatible with the flow-cytometer and their GFP and mCherry fluorescence were analysed. The GFP expression of each well was normalized by the mCherry expression that is not control by a riboswitch to decrease measurement error.

5.3.4. Flow cytometry

All cytometry measurements were performed on a CytoFlex S instrument from Beckman Coulter equipped with a 488 nm laser and a 561 nm laser for excitation of respectively GFP and mCherry. Emission light was bandpass-filtered at 510-520 nm of 610-620 nm respectively. Before measurement cells were diluted tenfold in PBS in a 96-well plate compatible with the device. For *in vivo* screening, 5000 events were recorded and for characterization in 24-well plates, 20000 events

were recorded. Cells were gated depending of their FSC, SSC in order to measure fluorescence of only viable, non-budding cells.

6. Material

The materials used in this thesis are listed as follow. All the chemicals were order with the highest purity possible. In table 6.1. are listed the chemicals and reagents, in table 6.2. the instrumentation, in table 6.3. the kits and commercially available systems, in table 6.4. the enzymes and proteins, in table 6.5. the cell strains used, in table 6.6. the consumables, in table 6.7. the buffer and solutions, in table 6.8. the oligonucleotide and in table 6.9. the plasmids.

All buffers and solutions were produced using deionized water or *aqua valde purificata*. If necessary, buffers and solutions were sterilized by autoclaving at 121°C and 2 bar for 20 min. Oligonucleotides were ordered from Sigma-Aldrich, Munich (desalted or RP1 purified). Cloned sequences were analysed by Seqlab, Göttingen using Sanger-Sequencing and capillary electrophoresis.

Table 6.1. List of chemicals and reagents

Chemicals and reagents	Manufacturer
[α-^{32}P] UTP [3000 Ci/mmol]	Hartmann Analytics, Braunschweig
Acetone	Roth, Karlsruhe
Acrylamid (Rotiphorese Gel 40)	Roth, Karlsruhe
Adenine	Roth, Karlsruhe
Agar	Oxoid, Heidelberg
Agarose peqGold Universal	Peqlab, Erlangen
Ammoniumperoxodisulfat	Roth, Karlsruhe
Ammonium sulfate	Roth, Karlsruhe
Amoxicillin	Roth, Karlsruhe
Ampicillin	Roth, Karlsruhe
rADP	Roth, Karlsruhe
dATP	Roth, Karlsruhe
rATP	Roth, Karlsruhe
Boric acid	Roth, Karlsruhe
Bromophenol blue	Roth, Karlsruhe
1-Butanol	Roth, Karlsruhe
Calcium chloride	Roth, Karlsruhe
Chloramphenicol	Roth, Karlsruhe
Citrate	Roth, Karlsruhe
Deoxynucleotide triphosphate (dNTP)	Peqlab, Erlangen
Dimethyl sulfoxide	Peqlab, Erlangen
Dulbecco's Phosphate buffered Saline (PBS)	Life Technologies, USA
Dithiothreitole (DTT)	Roth, Karlsruhe
Dynabeads M270 Streptavidin	Thermo Fisher, USA
Ethanol absolute, p.a.	Merck, Darmstadt
Ethanol, denatured	VWR, Darmstadt

Ethidium bromide	Roth, Karlsruhe
Ethylenediaminetetraacetic acid (EDTA)	Roth, Karlsruhe
FlowClean Cleaning Agent	Beckman Coulter
Fructose bi-phosphate (FBP)	-
Glucose, water-free	Roth, Karlsruhe
Glycerol, p.a.	Roth, Karlsruhe
Glyphosate	Sigma Aldrich
Hydrochloric acid (HCl)	Roth, Karlsruhe
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	-
Isopropanol, p.a.	VWR, Darmstadt
Kanamycin sulfate	Roth, Karlsruhe
MEM Amino acids, 50X	Sigma Aldrich
Magnesium chloride	Roth, Karlsruhe
Neomycin Trisulfate	Roth, Karlsruhe
Paromomycin sulfate	Roth, Karlsruhe
Polyacrylamid (Rotiphorese Gel 40, 19:1)	Roth, Karlsruhe
Potassium chloride (KCl)	Roth, Karlsruhe
Sodium chloride (NaCl)	Roth, Karlsruhe
Sodium dodecyl sulfate (SDS), pellets	Roth, Karlsruhe
Spermidine	Roth, Karlsruhe
TEMED	Roth, Karlsruhe
Tobramycin	Roth, Karlsruhe
Tris	Roth, Karlsruhe
Tryptone	Oxoid, Heidelberg
Triton-X-100	Roth, Karlsruhe
Tween 20	Roth, Karlsruhe
Urea	Roth, Karlsruhe
Viomycin	-
Xylene cyanole	Roth, Karlsruhe
Yeast extract	Oxoid, Heidelberg
Yeast nitrogen base (w/o ammonium sulfate)	Difco
Yeast synthetic drop-out (-Ura/Leu/Trp)	Sigma Aldrich

Table 6.2. List of utilized instruments

Instrument	Manufacturer
Accuracy weighing machine	Acculab, USA
Centrifuges	Heraeus Christ, Osterode
CytoFlex S (flow cytometer)	Beckman-Coulter, Krefeld
Electroporator MicroPulser™	Bio-Rad, Munich
Flourescence Stereo Microscope	Leica Microsystems, Wetzlar
Fluorolog FL3-22	Horiba, Darmstadt
Gel documentation with UV screen	INTAS, Göttingen
Heating block	VWR, Darmstadt
Incubation Shaker Multitron	Infors AG, Bottmingen
Incubator	Heraeus Christ, Osterode
Infinite® M200 plate reader	Tecan Trading AG, Switzerland
Magnetic stirrer IKA RET basic	IKA, Staufen
Milli-Q® water purification system with RNase filter	EMD Millipore, France
NanoDrop® ND-1000 spectral photometer	Peqlab, Erlangen
Scale	Acculab, USA
T100™ Thermal Cycler	Bio-Rad, Munich
Titramax 100, 1000 (platform shakers)	Heidolph, Schwabach
Inkubator 1000	Heidolph, Schwabach
Thermomixer comfort	Eppendorf AG, Hamburg

Table 6.3. List of utilized instruments

Kits and commercially available systems	Manufacturer
Frozen-EZ Yeast Transformation II	Zymo Research, USA
QIAquick Gel Extraction	QIAGEN, Hilden
QIAprep® Spin Miniprep Kit	QIAGEN, Hilden
QIAquick PCR Purification Kit	QIAGEN, Hilden

Table 6.4. List of utilized enzymes and ladders. Buffer were used as supplied and recommended by the manufacturer unless otherwise stated.

Enzymes and proteins	Manufacturer
Q5 High-Fidelity DNA Polymerase [2 U/µl]	New England Biolabs, USA
Antartic Phosphatase [1 U/µL]	New England Biolabs, USA
T4 DNA Ligase [400 U/µl]	New England Biolabs, USA
T4 Polynucleotide kinase [10 U/µl]	New England Biolabs, USA
Taq DNA Polymerase [5 U/µl]	New England Biolabs, USA
Taq DNA Ligase [40 U/µl]	New England Biolabs, USA

Restriction endonucleases	
AgeI-HF [20 U/µl]	New England Biolabs, USA
BbvCI [2 U/µl]	New England Biolabs, USA
BsiWI-HF [20 U/µl]	New England Biolabs, USA
BstXI [2 U/µl]	New England Biolabs, USA
ClaI [10 U/µl]	New England Biolabs, USA
NheI-HF [20 U/µl]	New England Biolabs, USA
SpeI-HF [20 U/µl]	New England Biolabs, USA
XbaI [20 U/µl]	New England Biolabs, USA

DNA/RNA Ladder	
peqGold Ultra Low Range DNA ladder II	PeqLab, Erlangen
peqGold 1 kB DNA ladder	PeqLab, Erlangen

Table 6.5. List of utilized cell strains

Cells	**Genotype/ Description**	**Reference**
Bacterial strains		
E. coli Top10	F- mcrA Δ(*mrr-hsd*RMS-*mcr*BC) Φ80*lac*ZΔM15 Δ *lac*X74 *rec*A1 *ara*D139 Δ(*araleu*)7697 *gal*U *gal*K *rps*L (StrR) *end*A1 *nup*G	Invitrogen, USA
Eukaryotic cell lines		
RS453α	*matα ade2-1 trp1-1 can1-100 leu2-3 his3-1 ura3-52*	Hillen Lab

Table 6.6. List of utilized consumables

Consumables	Manufacturer
Cellstar® cell culture plates, 12-well/ 24-well	Greiner Bio-One, Austria
Cuvettes, acrylic	Sarstedt, Nürnbrecht
Dispenser tips	Greiner, Nürtingen
Micro Pulser Electroporation Cuvettes, 0.1 cm gap	Bio-Rad, Munich
Breathseal, gas permeable, sterile	Greiner Bio-One, Austria
96-well U bottom/ flat bottom	Greiner Bio-One, Austria
Nitrile Gloves	Starlab, Hamburg
Petri dishes, plastic	Greiner, Nürtingen
Pipette tips, plastic	Greiner, Nürtingen Starlab, Hamburg Nerbe Plus, Winsen
Pipettes, plastic	Nerbe Plus, Winsen
Polyethylene tubes	Sarstedt, Nürnbrecht
Reaction tubes (1.5 ml, 2 ml)	Greiner, Nürtingen
Syringes	Becton Dickinson, Heidelberg

Table 6.7. List of utilized buffers and solutions

Buffer/ solution	Ingredients	Concentration
5x ISO buffer	Tris-HCl pH 7.5	0.5 M
	MgCl2	50 mM
	Each dNTP	1 mM
	DTT	50 mM
	PEG-8000	25% (w/v)
	NAD	5 mM
STET buffer	Sucrose	8% (w/v)
	Tris-HCl, pH 8	50 mM
	EDTA	50 mM
	Triton-X-100	5% (v/v)
YPD	Yeast extract	1% (w/v)
	Peptone	2% (w/v)
	Glucose	2% (w/v)
	Agar for plates	1.8% (w/v)

SCD -Ura	50x MEM Amino acids	1x
	Adenine	12 µg/ml
	Ammonium sulfate	0.55% (w/v)
	YNB	0.2% (w/v)
	Glucose	2% (w/v)
	Agar for plates	1.8% (w/v)
SCD –Ura/Leu/Trp	Yeast synthetic drop-out (-	0.15% (w/v)
	Ura/Leu/Trp)	0.55% (w/v)
	Ammonium sulfate	0.2% (w/v)
	YNB	2% (w/v)
	Glucose	0.076 mg/ml
	Uracil, if needed	0.38 mg/ml
	Leucine, if needed	0.076 mg/ml
	Tryptophan, if needed	1.8% (w/v)
	Agar for plates	
50x TAE	Tris	2 M
	Acetic acid	1 M
	EDTA	50 mM
	pH 8.3	
6x DNA loading dye	Tris-HCl pH 7.6	40 mM
	EDTA	1 mM
	Acetic acid	20 mM
	Glycerol	50% (v/v)
	Bromophenol blue	Spatula point
	Xylene cyanole	Spatula point
Ampicillin, stock solution	Ampicillin	100 mg/ml
	In 70% (v/v) EtOH	
LB medium	Tryptone	1% (w/v)
LB-Amp plates	Yeast extract	0.5% (w/v)
	NaCl	1% (w/v)
	Agar	2% (w/v)
	Ampicillin	100 µg/ml
SOC-medium	Yeast Extract	0.5%
	Tryptone	0.2%
	NaCl	10 mM
	KCl	2.5 mM
	MgCl2	10 mM
	MgSO4	10 mM
	Glucose	20 mM

Table 6.8. List of utilized oligonucleotides

Primer name	Sequence 5' --> 3'
Fwd1 Loop x2 CGOF11 2. part1	TATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCCACATACCACCAGAGACAGTC TTCTATTGGCTTC
Rev1 Loop x2 CGOF11 2 part 1	TCTGGTGGTATGTGGTTTGTTGTTGTTGTTGCCCTGCTCAGTAGAAGCCAATAGAAGACTGT CTCTGGTGGTATGTGGACCGGTAGATC
Fwd2 Loop x2 CGOF11 2 part2	CAACAACAACAAACCACATACCACCAGAGACAGTCTTCTATTGGCTTCTACTGAGCAGGGGC TAGCAAAGGAGAAGAAC
Rev2 Loop x2 CGOF11 2 part 2	GAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCCCTGCTCAGTAGAAGC
4xRev Ass Loop + GCx4	GGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCGCGCCCTGCTCAGCTGGT GGCGCGTATG
4xFwd Ass Loop + GCx4	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCAACATACGCGCCACCA GCTGAGCAGGGCGCGAAAATG
2xRev Ass Loop + GCx2	CAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCGCCCTGCTCAGCTGGTGGCGTA TGTTGACCGG
2xFwd Ass Loop + GCx2	GCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCAACATACGCCACCAGCTGA GCAGGGCG
3xRev Ass Loop + GCx3	GGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTGCGCCCTGCTCAGCTGGT GGCGC
3xFwd Ass Loop + GCx3	CAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCAACATAGCGCCACCAGCTGAGCAGG GCGCAAAATG
Rev Ass Bulgex2	GGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTCAGCCCTGCTCA GTTTGCTGGTGGCTGGTGGTATGTTG
Fwd Ass Bulgex2	CTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCAACATACCACCAGCCACCAG CAAACTGAGCAGGGCTG
Rev2 CGOF11 x2 GFP part 2	CAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCCCTGCTCAGTAGAAG CCAATAGAAGACTGTCTCTG
Fwd1 CGOF11 x2 GFP part 1	CTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTAGAACCCACAGTTCTACCACA TACCACCAGAGACAGTCTTCTATTGGC
Rev1 CGOF11 x2 GFP part 1	TGGTAGAACTGTGGGTTCTACCGGTTTGTTGTTGTTGTTGCCCTGCTCAGTAGAAGCCAATA GAAGACTGTCTCTGGTGGTATGTGGTAG
Fwd2 CGOF11 x2 GFP part 2	CAACAACAAACCGGTAGAACCCACAGTTCTACCACATACCACCAGAGACAGTCTTCTATTGG CTTCTACTGAGCAGGGGC
Fwd Pri CapSELEX 3.0	CCAAGTAATACGACTCACTATAGGGCAACTCCAAGCTAGATCTACCGG T
Rev Pri CapSELEX 3.0	AGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTT
Capture Oligo SEL EX AB 3.0	TAGAAGCCAGTAG[BtnTg]
CGOF1 fwd	ATCTACCGGTAGAACCCGCGGTTCTACAACATACCACCAGAGACAGCC ATCTACTGGCTT
CGOF1 rev	CTTTGCTAGCCATTTTCCCTGCTCAGTAGAAGCCAGTAGATGGCTGTC TCTGGTGGTAT
CGOF2 fwd	ATCTACCGGTAGAACCCACAGTTCTACAACATACCACCAGAGACAGCC TTCTACTGGCTT
CGOF2 rev	CTTTGCTAGCCATTTTCCCTGCTCAGTAGAAGCCAGTAGAAGGCTGTC TCTGGTGGTAT
CGOF3 fwd	ATCTACCGGTAGAACCCACAGTTCTACAACATACCACCAGAGACAGCC ATCTACTGGCTT
CGOF3 rev	CTTTGCTAGCCATTTTCCCTGCTCAGTAGAAGCCAGTAGATGGCTGTC TCTGGTGGTAT
ADH fwd	GCACAATATTTCAAGCTATACC
GFP rev	CCACTGACAGAAAATTTGTGCCC
Adh1term fwd (Bsi Wi RS) AB	CTCAGCGTACGAGCGACCTCATGCTATACCTGAGAAAGC
mCherry-switch rev	GTCTTCTACTGGCTTCTACTGAGCAGGGAAAATGGTGAGCAAGGGCGAGGAGG
mCherry switch fwd	CCTCCTCGCCCTTGCTCACCATTTTCCCTGCTCAGTAGAAGCCAGTAGAAGAC
switch-TEF prom rev	GAAAGAAAGCATAGCAATCTAATCTAAGTTTTCTAGACTATACCAAGCATACAATCAACTCCA AGC
Switch-TEF prom fwd	GCTTGGAGTTGATTGTATGCTTGGTATAGTCTAGAAAACTTAGATTAGATTGCTATGCTTTCTT TC
TEFprom rev	CATGCGACTAGTGGCCTATGCATAGCTTCGAAACG

CO 3.0 NNN +2 AB	NTAGAAGCCAGTAGN[BtnTg]
CO 3.0 +4 AB	NNTAGAAGCCAGTAGNN[BtnTg]
CO 3.0 +6 AB	NNNTAGAAGCCAGTAGNNN[BtnTg]
CGOF4 fwd	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCATACCACCAGAGACAG CCTTCTACGGGCGTCTACTGAGCAGGG
CGOF4-5 rev	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGACGCCCGTAGAAGGCTGTCTCTGG
CGOF5 fwd	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCAAACCACCAGAGACA GCCTTCTACGGGCGTCTACTGAGCAGGG
CGOF6 fwd	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCCACCAGAGACAGCCT TCTACTGGCTTCTACTGAGCAGGG
CGOF6 rev	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGAAGCCAGTAGAAGGCTGTC
CGOF7 fwd	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCATACCACCAGAGAGA AATCTACTGAGCAGGG
CGOF7 rev	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGATTTCTCTCTGGTGGTATG
Fwd A45 C	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTAGAACCCACAGTTCTAC CACAAACCACCAGAGACAGTCTT
Rev A45C	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGAAGCCAGTAGAAGACTGTCTCTGGTGGTTTGTGGTAG
Fwd U76G U72G C68 U C71U	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTAGAACCCACAGTTCTAC AACAAACCACCAGAGACAGTCTT
Rev U76G	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGACGCCAGTAGAAGACTGTCTCTGGTGGTTTGTTGTAG
Rev U72G	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGAAGCCCGTAGAAGACTGTCTCTGGTGGTTTGTTGTAG
Rev C68U	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGAAGCCAGTAAAAGACTGTCTCTGGTGGTTTGTTGTAG
Rev C71U	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGAAGCCAATAGAAGACTGTCTCTGGTGGTTTGTTGTAG
Fwd C36G	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTAGAACCCAGAGTTCTAC AACAAACCACCAGAGACAGTCTT
Rev C36G	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGAAGCCAGTAGAAGACTGTCTCTGGTGGTTTGTTGTAG
CGOF8 fwd	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTCAACAACAACAAACAA CTACAACATACCACCAGAGACAGCCATC
CGOF8 rev	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGAAGCCAGTAGATGGCTGTCTCTGGTGGTATGTTG
Fwd C34U	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTAGAACCTACAGTTCTAC AACAAACCACCAGAGACAGTCTT
Fwd A49U	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTAGAACCCACAGTTCTAC AACATACCACCAGAGACAGTCTT
Rev A49U	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGAAGCCAGTAGAAGACTGTCTCTGGTGGTATGTTGTAG
PARO-RBS fwd	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCGGTAGAACCCACAGTTCTAC CACATACCACCAGAGACAGTCTT
PARO-RBS rev	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGAAGCCAATAGAAGACTGTCTCTGGTGGTATGTGGTAG
Fwd Neo Paro Tandem	CAAGCTATACCAAGCATACAATCAACTCCAAGCTAGATCTACCCGGCATAGCTTGTCCTTTAAT GGTCCTATGTCGACATACCACCAGAGACAGTCTTC
Rev Neo Paro Tandem	CAACAAGAATTGGGACAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTC AGTAGAAGCCAATAGAAGACTGTCTCTGGTGGTATGTCG
Fwd Paro Neo Tandem	CCAAGCATACAATCAACTCCAAGCTAGATCTACCACATACCACCAGAGACAGTCTTCTATTGG CTTCTACTGAGCAGGGCAACAAC
Rev Paro Neo Tandem	CAACTCCAGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCGACATAGGACCATTAAAGGACA AGCTATGCCGTTGTTGTTGCCCTGCTCAGTAGAAGCC

Paro RBS cherry rev	CCACCAGAGACAGTCTTCTATTGGCTTCTACTGAGCAGGGAAAATGGTGAGCAAGGGCGAGGAGG
Paro RBS TEF fwd	GCCAATAGAAGACTGTCTCTGGTGGTATGTGGTAGAACTGTGGGTTCTACCGGTTCTAGAAAACTTAGATTAGATTGCTATGCTTTC
Fwd pri pool_2 AB	TAATACGACTCACTATAGTGGATCCGACCGTGTATGT
Rev pri pool_2 AB	CTGACGTCCATCCGCCTGTT
COP_2	TAGAAGTTAGTA[BtnTg]
Fwd Paro switch CAA RNA	CCAAGTAATACGACTCACTATAGGGCAACTCCAAGCTAGATCTACCGGTACAACAACAACAACAACCACATACCACCAGAGACAGTCTTC
Rev Paro switch CAA RNA	AGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTCAGTAGAAGCCAATAGAAGACTGTCTCTGGTGGTATGTGGTTG
Fwd Paro switch RNA	CCAAGTAATACGACTCACTATAGGGCAACTCCAAGCTAGATCTACCGGTAGAACCCACAGTTCTACCACATACCACCAGAGACAGTCTTC
Rev Paro switch RNA	AGTGAAAAGTTCTTCTCCTTTGCTAGCCATTTTCCCTGCTCAGTAGAAGCCAATAGAAGACTGTCTCTGGTGGTATGTGGTAGAAC
Fwd 1 Switch Prom paro	GGTTTGTTGTAGAACTGTGGGTTCTACCGGTAGATCTAGCTTGGAGTTGATTGTATGCTTGGTATAGTCTAGAAAACTTAGATTAGATTGCTATGCTTTC
Rev 1 Prom Cla I paro	GAAAATTCTTTTTTTGATTTTTTTCTCTTTCGATGACCTCCCATCGATATTTAAGTTAATAAACGGCCTTCAATTTCCC
Fwd 2 Cherry Bbv paro	CGCCGGGCAGCTGCACGGGCTTCTTGGCCTTGTAGGTGGTCTTGACCTCAGCGTCGTAGTGGCCGCCGTCCTTCAGCTTC
Rev 2 Cherry switch paro	CCGGTAGAACCCACAGTTCTACAACAAACCACCAGAGACAGTCTTCTACTGGCTTCTACTGAGCAGGGAAAATGGTGAGCAAGGGCGAGGAGGATAAC
Fwd 3 Neo Cherry	CGGCATAGCTTGTCCTTTAATGGTCCTATGTCGAAAATGGTGAGCAAGGGCGAGGAGGATAACATGGCC
Rev 3 Cherry end BbV Neo	GCGGCCCTCGGCGCGTTCGTACTGTTCCAC
Fwd 4 TEF prom Cla Neo	GAGACCGCCTCGTTTCTTTTTCTTCGTCGAAAAAGGC
Rev 4 Neo mcherry	CGACATAGGACCATTAAAGGACAAGCTATGCCGTCTAGAAAACTTAGATTAGATTGCTATGCTTTCTTTC

Table 6.9. List of utilized plasmid

Plasmid name	Description
pWHE601	Adapted from Dr. Julia Weigand (91)
pWHE601*	Adapted from Dr. Julia Weigand (91)
Neo-Apt	Adapted from Dr. Julia Weigand, M4 riboswitch (91)
pCBB05	Centromeric plasmid with GFP and mCherry gene, adapted from Dr. Cristina Bofill-Bosch
pCBB06	Deletion of start codon of pCBB05, adapted from Dr. Cristine Bofill-Bosch
pCBB07	Insertion of the M4 riboswitch (from Dr. Julia Weigand work) in front of GFP, adapted from Dr. Cristine Bofill-Bosch
pCBB06-PARO	Insertion of the PARO riboswitch in the pCBB06 in front of GFP
Neo-Paro	Insertion of the tandem Neo-Paro riboswitch in the pCBB06 in front of GFP

6.1. pWHE601, pWHE601*

Vector pWHE601 (90) was used as a positive control for *in vivo* screening assays for the splicing regulation project. For selection in bacteria a β-lactamase gene is encoded conferring resistance against ampicillin. To select for the plasmid in yeast *URA3* is expressed that codes for orotidine-5'-phosphate (OMP) decarboxylase enabling Ura- strains to grow on synthetic media without uracil. The vector contains a 2µ origin of replication (ori) for its replication in yeast cells. The expression cassette is comprised of an *ADH1* promoter and terminator and a *GFP+* reporter gene that is 5'-scarred with the *NheI* restriction site. Vector pWHE601* represents a derivative of pWHE601 with the following minor modifications. (i) The *AflII* restriction site was exchanged for the restriction site for *AgeI,* (ii) the start codon was deleted and substituted by CTCTTC now spacing *AgeI* and *NheI.* This vector was used as a negative control in *in vivo* screening assays for the splicing regulation project and as a basis for introducing all aptamers from the PAC SELEX.

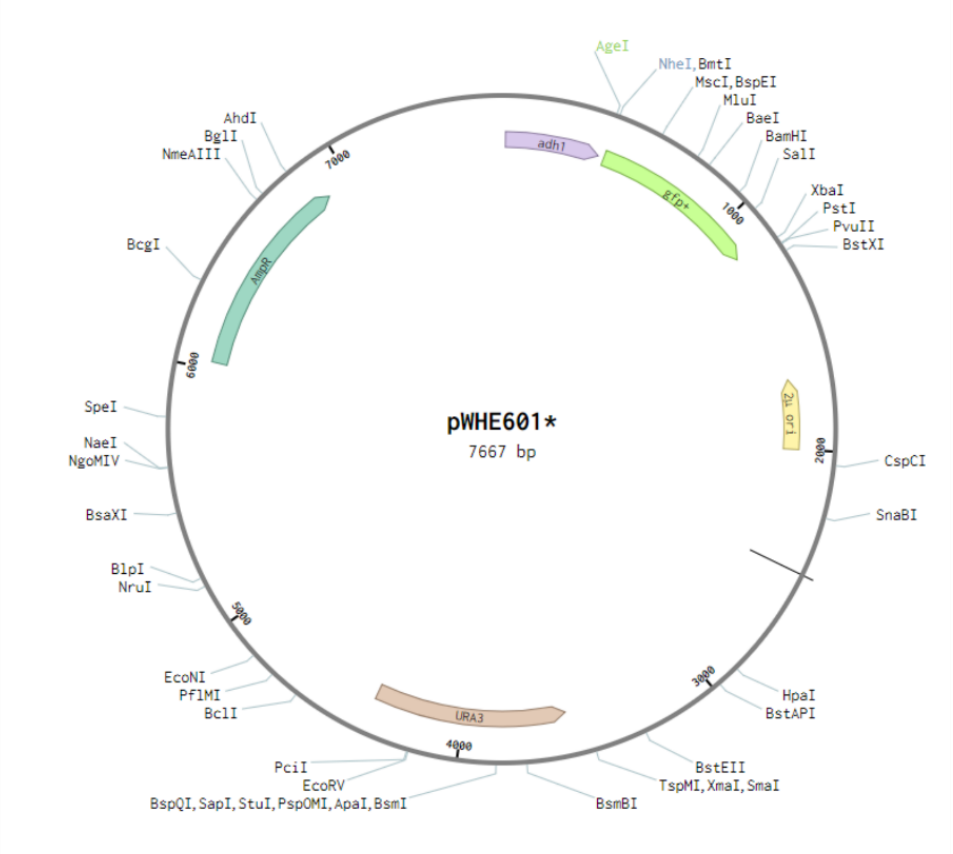

Figure 6.1. The plasmid map shows schematically the construction of the pWHE601* plasmid used for in vivo screening for splicing regulation. The plasmid possesses a GFP gene (green) and under the control of the ADH promotor (blue) by the digestion of the plasmid by AgeI (3317) and NheI (3323). The gene URA3 (yellow) is the selection marker in yeast coding for orotidine 5'-phosphate decarboxylase that allows growth on synthetic media without uracil. The ampicillin resistance gene (AmpR) was used as selection marker for E. coli. The plasmid possesses also a 2µ origin of replication sequence.

6.2. pCBB05, pCBB06, pCBB07

The plasmid pCBB06 is derived from the pWHE601*. The GFP domain (including promotor and terminator), ampicillin resistance and URA3 gene were kept. The plasmid map (figure 6.2.) shows schematically the construction of the pCBB06 plasmid used for in vivo screening. However, the 2µ replication sequence was exchanged to a CEN6/ARS4 sequence to make it a centromeric plasmid. In addition, the plasmid contains the mCherry gene under the control of a TEF promoter used later on for GFP expression normalization. Plasmid pCBB05 is identical to pCBB06 except the addition of the Kozak-start sequence in front of the GFP gene. This plasmid was used as a positive control in the *in vivo* screening assays. For switching control, pCBB07 was created where the aptamer M4 (91) for neomycin was inserted after the Age-I restriction site of pCBB05.

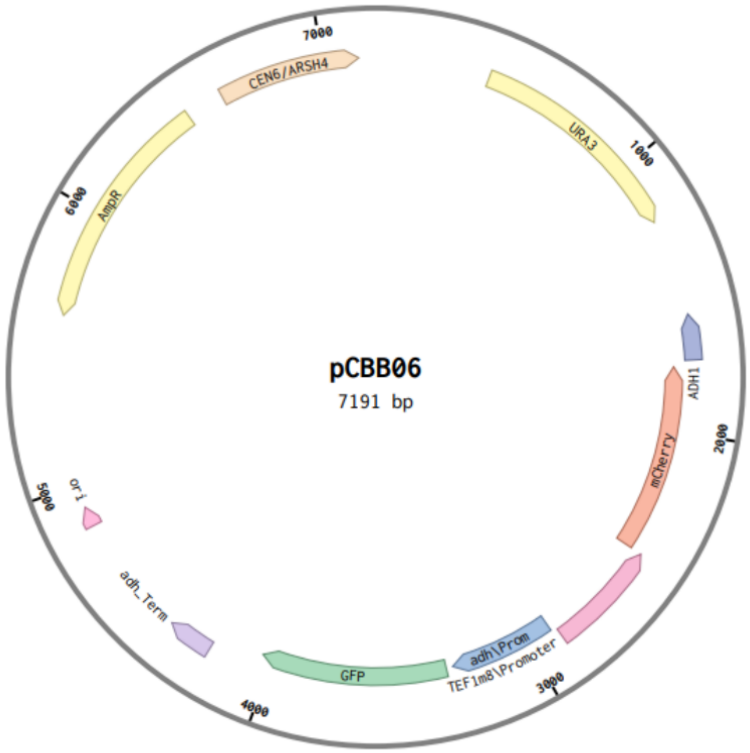

Figure 6.2. The plasmid map shows schematically the construction of the pCBB06 plasmid used for in vivo screening for translation initiation regulation. The plasmid possesses a GFP gene (green) and under the control of the ADH promotor (blue) by the digestion of the plasmid by AgeI (3317) and NheI (3323). GFP expression was normalized by the mCherry expression (orange) controlled by a TEF1m8 promotor. The gene URA3 (yellow) is the selection marker in yeast coding for orotidine 5'-phosphate decarboxylase that allows growth on synthetic media without uracil. The ampicillin resistance gene (AmpR) was used as selection marker for E. coli.

7. Appendix

7.1. Abbreviations

Table 7.1. List of abbreviations

Abbreviation		Abbreviation	
% (v/v)	% (volume/volume)	fwd	forward
% (w/v)	% (weight/volume)	LB	lysogeny broth
ADH1	Alcohol dehydrogenase 1	mRNA	messenger RNA
Amp	ampicillin	nt	nucleotides
bp	base pair	PBS	phosphate buffered saline
DMSO	Dimethylsulfoxide	RNA	ribonucleic acid
dNTP	deoxynucleoside triphosphate	SDS	Sodium dodecyl sulfate
et al.	*et alii*	TEF1	Transcription elongation factor EF-1 alpha
EtOH	Ethanol	UTR	untranslated region

7.2. Units

Table 7.2. List of units

Units	
°C	degree Celsius
Da	Dalton
g	gram
h	hour
L	liter
M	molar
min	minutes
rpm	rounds per minute

sec	second
U	unit
V	volt
v/v	volume per volume
w/v	weight per volume

7.3. Prefixes

Table 7.3. List of used prefixes

Dimensions	
M	mega (10^6)
K	kilo (10^3)
M	milli (10^{-3})
µ	micro (10^{-6})
N	nano (10^{-9})
P	pico (10^{-12})
F	femto (10^{-15})

7.4. Nucleobases

Table 7.4. Abbreviations for nucleobases

Dimensions	
A	adenine
C	cytosine
G	guanine
T	thymine
U	uracil
N	A, C, G, T, U

8. References

1. Wellhausen,R. and Oye,K.A. (2008) Intellectual property and the commons in synthetic biology: Strategies to facilitate an emerging technology. *2007 Atlanta Conf. Sci. Technol. Innov. Policy, ACSTIP*, 10.1109/ACSTIP.2007.4472901.
2. Vriend,H. and Walhout,A. (2006) Constructing Life-Early social reflections on the emerging field of synthetic biology.
3. Bandyopadhyay,S., Mehta,M., Kuo,D., Sung,M., Chuang,R., Fiedler,D., Dutkowski,J., Guénolé,A. and Attikum,H. Van (2010) References and Notes 1. **329**, 1385–1389.
4. Qi,L.S., Larson,M.H., Gilbert,L.A., Doudna,J.A., Weissman,J.S., Arkin,A.P. and Lim,W.A. (2013) Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell*, **152**, 1173–1183.
5. Gibson,D.G., Benders,G.A., Andrews-pfannkoch,C., Denisova,E.A., Baden-tillson,H., Zaveri,J., Stockwell,T.B., Brownley,A., Thomas,D.W., Algire,M.A., *et al.* (2008) and Cloning of a Mycoplasma genitalium Genome. *Science (80-.).*, **319**, 1215–1221.
6. Dymond,J.S., Richardson,S.M., Coombes,C.E., Babatz,T., Muller,H., Annaluru,N., Blake,W.J., Schwerzmann,J.W., Dai,J., Lindstrom,D.L., *et al.* (2011) Synthetic chromosome arms function in yeast and generate phenotypic diversity by design. *Nature*, **477**, 471–476.
7. Schneider,C., Bronstein,L., Diemer,J., Koeppl,H. and Suess,B. (2017) ROC'n'Ribo: Characterizing a Riboswitching Expression System by Modeling Single-Cell Data. *ACS Synth. Biol.*, **6**, 1211–1224.
8. Groher,F., Bofill-Bosch,C., Schneider,C., Braun,J., Jager,S., Geißler,K., Hamacher,K. and Suess,B. (2018) Riboswitching with ciprofloxacin-development and characterization of a novel RNA regulator. *Nucleic Acids Res.*, **46**, 2121–2132.
9. Mathews,D.H., Turner,D.H. and Zuker,M. (2007) RNA Secondary Structure Prediction. *Curr. Protoc. Nucleic Acid Chem.*, 10.1002/0471142700.nc1102s28.
10. Zuker,M. (1989) On finding all suboptimal foldings of an RNA molecule. *Science (80-.).*, **244**, 48 LP-52.
11. Batey,R.T., Rambo,R.P. and Doudna,J.A. (1999) Tertiary motifs in RNA structure and folding. *Angew. Chemie - Int. Ed.*, **38**, 2326–2343.
12. Steiner,T. (2002) The hydrogen bond in the solid state. *Angew. Chem. Int. Ed.*, **41**, 49–76.
13. Lusk,J.E., Williams,R.J.P. and Kennedy,E.P. (1968) Magnesium and the growth of escherichia coli. *J. Biol. Chem.*, **243**, 2618–2624.
14. Olubunmi Bolanle,A. and Oluwaseyi Israel,M. (2014) Effect of Sesamum indicum L. seed oil supplementation on hepatic and renal mineral concentrations of hypercholesterolemic rats. *Am. J. Life Sci.*, **2**, 308–311.
15. Reuss,A.J., Vogel,M., Weigand,J.E., Suess,B. and Wachtveitl,J. (2014) Tetracycline determines the conformation of its aptamer at physiological magnesium concentrations. *Biophys. J.*, **107**, 2953–62.
16. Wilson,K.A., Holland,D.J. and Wetmore,S.D. (2016) Topology of RNA – protein nucleobase – amino acid π – π interactions and comparison to analogous DNA – protein π – π contacts. 10.1261/rna.054924.115.696.
17. Desfrançois,C., Carles,S. and Schermann,J.P. (2000) Weakly bound clusters of biological interest. *Chem. Rev.*, **100**, 3943–3962.
18. Gorin,A.A., Zhurkin,V.B. and Wilma K. (1995) B-DNA Twisting Correlates with Base-pair Morphology. *J. Mol. Biol.*, **247**, 34–48.
19. Olson,W.K., Gorin,A.A., Lu,X.-J., Hock,L.M. and Zhurkin,V.B. (1998) DNA sequence-dependent deformability deduced from protein–DNA crystal complexes. *Proc. Natl. Acad. Sci.*, **95**, 11163 LP-11168.
20. Manalo,M.N., Pérez,L.M. and LiWang,A. (2007) Hydrogen-bonding and π-π base-stacking

interactions are coupled in DNA, as suggested by calculated and experimental trans-hbond deuterium isotope shifts. *J. Am. Chem. Soc.*, **129**, 11298–11299.
21. Matta,C.F., Castillo,N. and Boyd,R.J. (2006) Extended weak bonding interactions in DNA: π-stacking (base-base), base-backbone, and backbone-backbone interactions. *J. Phys. Chem. B*, **110**, 563–578.
22. Ok,T., Lee,J. and Lee,H. (2011) γ 3 PNA : Peptide Analogue of Glycol Nucleic Acid †. **32**, 2863–2864.
23. Mignon,P., Loverix,S., Steyaert,J. and Geerlings,P. (2005) Influence of the π-π interaction on the hydrogen bonding capacity of stacked DNA/RNA bases. *Nucleic Acids Res.*, **33**, 1779–1789.
24. McKeague,M. and Derosa,M.C. (2012) Challenges and opportunities for small molecule aptamer development. *J. Nucleic Acids*, **2012**.
25. Birch,J.R. and Racher,A.J. (2006) Antibody production. *Adv. Drug Deliv. Rev.*, **58**, 671–685.
26. Thirunavukarasu,D., Chen,T., Liu,Z., Hongdilokkul,N. and Romesberg,F.E. (2017) Selection of 2'-Fluoro-Modified Aptamers with Optimized Properties. *J. Am. Chem. Soc.*, **139**, 2892–2895.
27. E. Wang,R., Zhang,Y., Cai,J., Cai,W. and Gao,T. (2011) Aptamer-Based Fluorescent Biosensors. *Curr. Med. Chem.*, **18**, 4175–4184.
28. Mascini,M. (2008) Aptamers and their applications. *Anal. Bioanal. Chem.*, **390**, 987–988.
29. Paper,W. (2013) Long-term storage of nucleic acids.
30. Ye,B.-F., Zhao,Y.-J., Cheng,Y., Li,T.-T., Xie,Z.-Y., Zhao,X.-W. and Gu,Z.-Z. (2012) Colorimetric photonic hydrogel aptasensor for the screening of heavy metal ions. *Nanoscale*, **4**, 5998–6003.
31. Meli,M., Vergne,J., Décout,J.-L. and Maurel,M.-C. (2002) Adenine-aptamer complexes: a bipartite RNA site that binds the adenine nucleic base. *J. Biol. Chem.*, **277**, 2104—2111.
32. Lauhon,C.T. and Szostak,J.W. (1995) RNA aptamers that bind flavin and nicotinamide redox cofactors. *J. Am. Chem. Soc.*, **117**, 1246–1257.
33. Stojanovic,M.N., de Prada,P. and Landry,D.W. (2001) Aptamer-based folding fluorescent sensor for cocaine. *J. Am. Chem. Soc.*, **123**, 4928–4931.
34. Stojanovic,M.N., de Prada,P. and Landry,D.W. (2000) Fluorescent Sensors Based on Aptamer Self-Assembly. *J. Am. Chem. Soc.*, **122**, 11547–11548.
35. Nutiu,R. and Li,Y. (2005) In vitro selection of structure-switching signaling aptamers. *Angew. Chemie - Int. Ed.*, **44**, 1061–1065.
36. Elowe,N.H., Nutiu,R., Allali-Hassani,A., Cechetto,J.D., Hughes,D.W., Li,Y. and Brown,E.D. (2006) Small-molecule screening made simple for a difficult target with a signaling nucleic acid aptamer that reports on deaminase activity. *Angew. Chemie - Int. Ed.*, **45**, 5648–5652.
37. Olek,M., Büsgen,T., Hilgendorff,M. and Giersig,M. (2006) Quantum Dot Modified Multiwall Carbon Nanotubes. *J. Phys. Chem. B*, **110**, 12901–12904.
38. Chhabra,R., Sharma,J., Wang,H., Zou,S., Lin,S., Yan,H., Lindsay,S. and Liu,Y. (2009) Distance-dependent interactions between gold nanoparticles and fluorescent molecules with DNA as tunable spacers. *Nanotechnology*, **20**.
39. Huang,Y., Zhao,S., Liang,H., Chen,Z.F. and Liu,Y.M. (2011) Multiplex detection of endonucleases by using a multicolor gold nanobeacon. *Chem. - A Eur. J.*, **17**, 7313–7319.
40. Zhao,W., Chiuman,W., Brook,M.A. and Li,Y. (2007) Simple and rapid colorimetric biosensors based on DNA aptamer and noncrosslinking gold nanoparticle aggregation. *ChemBioChem*, **8**, 727–731.
41. Liu,J. and Lu,Y. (2005) Fast colorimetric sensing of adenosine and cocaine based on a general sensor design involving aptamers and nanoparticles. *Angew. Chemie - Int. Ed.*, **45**, 90–94.
42. Tombelli,S., Minunni,M., Luzi,E. and Mascini,M. (2005) Aptamer-based biosensors for the detection of HIV-1 Tat protein. *Bioelectrochemistry*, **67**, 135–141.
43. Luzi,E., Minunni,M., Tombelli,S. and Mascini,M. (2003) New trends in affinity sensing: Aptamers for ligand binding. *TrAC - Trends Anal. Chem.*, **22**, 810–818.

44. Song,S., Wang,L., Li,J., Fan,C. and Zhao,J. (2008) Aptamer-based biosensors. *TrAC - Trends Anal. Chem.*, **27**, 108–117.
45. Wang,L., Zhu,C., Han,L., Jin,L., Zhou,M. and Dong,S. (2011) Label-free, regenerative and sensitive surface plasmon resonance and electrochemical aptasensors based on graphene. *Chem. Commun.*, **47**, 7794–7796.
46. Vivekananda,J. and Kiel,J.L. (2006) Anti-Francisella tularensis DNA aptamers detect tularemia antigen from different subspecies by Aptamer-Linked Immobilized Sorbent Assay. *Lab. Investig.*, **86**, 610.
47. Paige,J.S., Wu,K.Y. and Jaffrey,S.R. (2011) RNA Mimics of Green Fluorescent Protein. *Science (80-.).*, **333**, 642–646.
48. Strack,R.L., Song,W. and Jaffrey,S.R. (2013) Using Spinach-based sensors for fluorescence imaging of intracellular metabolites and proteins in living bacteria. *Nat. Protoc.*, **9**, 146–155.
49. DasGupta,S., Shelke,S. a., Li,N. and Piccirilli,J. a. (2015) Spinach RNA aptamer detects lead(<scp>ii</scp>) with high selectivity. *Chem. Commun.*, **51**, 9034–9037.
50. Pothoulakis,G., Ceroni,F., Reeve,B. and Ellis,T. (2013) The Spinach RNA aptamer as a characterisation tool for synthetic biology. *ACS Synth. Biol.*, 10.1021/sb400089c.
51. Lee,J.H., Jucker,F. and Pardi,A. (2008) Imino proton exchange rates imply an induced-fit binding mechanism for the VEGF165-targeting aptamer, Macugen. *FEBS Lett.*, **582**, 1835–1839.
52. Bates,P.J., Laber,D.A., Miller,D.M., Thomas,S.D. and Trent,J.O. (2009) Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. *Exp. Mol. Pathol.*, **86**, 151–164.
53. Diener,J.L., Daniel Lagassé,H.A., Duerschmied,D., Merhi,Y., Tanguay,J.F., Hutabarat,R., Gilbert,J., Wagner,D.D. and Schaub,R.G. (2009) Inhibition of von Willebrand factor-mediated platelet activation and thrombosis by the anti-von Willebrand factor A1-domain aptamer ARC1779. *J. Thromb. Haemost.*, **7**, 1155–1162.
54. Tasset,D.M., Kubik,M.F. and Steiner,W. (1997) Oligonucleotide inhibitors of human thrombin that bind distinct epitopes. *J. Mol. Biol.*, **272**, 688–698.
55. Min,K., Jo,H., Song,K., Cho,M., Chun,Y.S., Jon,S., Kim,W.J. and Ban,C. (2011) Dual-aptamer-based delivery vehicle of doxorubicin to both PSMA (+) and PSMA (-) prostate cancers. *Biomaterials*, **32**, 2124–2132.
56. Pollard,J., Bell,S.D. and Ellington,A.D. (2001) Design, Synthesis, and Amplification of DNA Pools for Construction of Combinatorial Pools and Libraries. *Curr. Protoc. Mol. Biol.*, 10.1002/0471142727.mb2402s52.
57. Hall,B., Arshad,S., Seo,K., Bowman,C., Corley,M., Jhaveri,S.D. and Ellington,A.D. In Vitro Selection of RNA Aptamers to a Protein Target by Filter Immobilization. *Curr. Protoc. Mol. Biol.*, **89**, 24.3.1-24.3.27.
58. Tuerk,C. and Gold,L. (1990) Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science*, **249**, 505–510.
59. Ellington,A.D. and Szostak,J.W. (1990) In vitro selection of RNA molecules that bind specific ligands. *Nature*, **346**, 818.
60. Robertson,D.L. and Joyce,G.F. (1990) Selection in vitro of an RNA enzyme that specifically cleaves single-stranded DNA. *Nature*, **344**, 467–468.
61. Geiger,A., Burgstaller,P., von der Eltz,H., Roeder,A. and Famulok,M. (1996) RNA aptamers that bind L-arginine with sub-micromolar dissociation constants and high enantioselectivity. *Nucleic Acids Res.*, **24**, 1029–1036.
62. Jenison,R.D., Gill,S.C., Pardi,A. and Polisky,B. (1994) High-resolution molecular discrimination by RNA. *Science (80-.).*, **263**, 1425–1429.
63. Rangel,A.E., Chen,Z., Ayele,T.M. and Heemstra,J.M. (2018) In vitro selection of an XNA aptamer capable of small-molecule recognition. *Nucleic Acids Res.*, 10.1093/nar/gky667.

64. Karlsen,K.K. and Wengel,J. (2012) Locked nucleic acid and aptamers. *Nucleic Acid Ther.*, **22**, 366–70.
65. Eulberg,D. and Klussmann,S. (2003) Spiegelmers: Biostable aptamers. *ChemBioChem*, **4**, 979–983.
66. Zhu,B., Hernandez,A., Tan,M., Wollenhaupt,J., Tabor,S. and Richardson,C.C. (2015) Synthesis of 2'-Fluoro RNA by Syn5 RNA polymerase. *Nucleic Acids Res.*, **43**, e94.
67. Shangguan,D., Li,Y., Tang,Z., Cao,Z.C., Chen,H.W., Mallikaratchy,P., Sefah,K., Yang,C.J. and Tan,W. (2006) Aptamers evolved from live cells as effective molecular probes for cancer study. *Proc. Natl. Acad. Sci.*, **103**, 11838 LP-11843.
68. Pestourie,C., Cerchia,L., Gombert,K., Aissouni,Y., Boulay,J., Franciscis,V. De, Libri,D., Tavitian,B. and Ducongé,F. (2006) Comparison of Different Strategies to Select Aptamers Against a Transmembrane Protein Target. *Oligonucleotides*, **16**, 323–335.
69. Nallani,M., Andreasson-Ochsner,M., Tan,C.-W.D., Sinner,E.-K., Wisantoso,Y., Geifman-Shochat,S. and Hunziker,W. (2011) Proteopolymersomes: In vitro production of a membrane protein in polymersome membranes. *Biointerphases*, **6**, 153–157.
70. Mendonsa,S.D. and Bowser,M.T. (2004) In vitro evolution of functional DNA using capillary electrophoresis. *J. Am. Chem. Soc.*, **126**, 20–21.
71. Stoltenburg,R., Nikolaus,N. and Strehlitz,B. (2012) Capture-SELEX : Selection of DNA Aptamers for Aminoglycoside Antibiotics. **2012**.
72. Wallis,M.G., von Ahsen,U., Schroeder,R. and Famulok,M. (1995) A novel RNA motif for neomycin recognition. *Chem. Biol.*, **2**, 543–552.
73. Cech,T.R. and Golden,B.L. (1999) Building a Catalytic Active Site Using Only RNA. *Mech. RNA Catal.*
74. Breaker,R.R. (1997) In Vitro Selection of Catalytic Polynucleotides. *Chem. Rev.*, **97**, 371–390.
75. Hermann,T. and Patel,D.J. (2000) Adaptive recognition by nucleic acid aptamers. *Science*, **287**, 820–825.
76. Osborne,S.E. and Ellington,A.D. (1997) Nucleic Acid Selection and the Challenge of Combinatorial Chemistry. *Chem. Rev.*, **97**, 349–370.
77. Webb,E. and Downs,D. (1997) Characterization of thiL, encoding thiamin-monophosphate kinase, in salmonella typhimurium. *J. Biol. Chem.*, **272**, 15702–15707.
78. Winkler,W., Nahvi,A. and Breaker,R.R. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*, **419**, 952–956.
79. Mironov,A.S., Gusarov,I., Rafikov,R., Lopez,L.E., Shatalin,K., Kreneva,R. a., Perumov,D. a. and Nudler,E. (2002) Sensing small molecules by nascent RNA: A mechanism to control transcription in bacteria. *Cell*, **111**, 747–756.
80. Serganov,A. and Nudler,E. (2013) A decade of riboswitches. *Cell*, **152**, 17–24.
81. Breaker,R.R. (2010) Riboswitches and the RNA World Riboswitches and the RNA World. 10.1101/cshperspect.a003566.
82. Suess,B., Fink,B., Berens,C., Stentz,R. and Hillen,W. (2004) A theophylline responsive riboswitch based on helix slipping controls gene expression in vivo. *Nucleic Acids Res.*, **32**, 1610–1614.
83. Zhao,R., Collins,E.J., Bourret,R.B. and Silversmith,R.E. (2002) Structure and catalytic mechanism of the e. coli chemotaxis phosphatase chez. *Nat. Struct. Biol.*, **9**, 570–575.
84. Seeliger,J.C., Topp,S., Sogi,K.M., Previti,M.L., Gallivan,J.P. and Bertozzi,C.R. (2012) A Riboswitch-Based Inducible Gene Expression System for Mycobacteria. *PLoS One*, **7**, 1–6.
85. Rudolph,M.M., Vockenhuber,M.P. and Suess,B. (2013) Synthetic riboswitches for the conditional control of gene expression in *Streptomyces coelicolor*. *Microbiology*, **159**, 1416–1422.
86. Chang,A.L., Wolf,J.J. and Smolke,C.D. (2012) Synthetic RNA switches as a tool for temporal and spatial control over gene expression. *Curr. Opin. Biotechnol.*, **23**, 679–688.
87. Nakahira,Y., Ogawa, a., Asano,H., Oyama,T. and Tozawa,Y. (2013) Theophylline-Dependent

Riboswitch as a Novel Genetic Tool for Strict Regulation of Protein Expression in Cyanobacterium Synechococcus elongatus PCC 7942. *Plant Cell Physiol.*, **54**, 1724–1735.
88. Busch,A. and Backofen,R. (2006) INFO-RNA—a fast approach to inverse RNA folding. *Bioinformatics*, **22**, 1823–1831.
89. Werstuck,G. and Green,M.R. (1998) Controlling Gene Expression in Living Cells Through Small Molecule-RNA Interactions. *Science (80-.).*, **282**, 296–298.
90. Suess,B., Hanson,S., Berens,C., Fink,B., Schroeder,R. and Hillen,W. (2003) Conditional gene expression by controlling translation with tetracycline-binding aptamers. *Nucleic Acids Res.*, **31**, 1853–1858.
91. Weigand,J.E., Sanchez,M., Gunnesch,E.-B., Zeiher,S., Schroeder,R. and Suess,B. (2008) Screening for engineered neomycin riboswitches that control translation initiation. *RNA*, **14**, 89–97.
92. Weigand,J.E. and Suess,B. (2007) Tetracycline aptamer-controlled regulation of pre-mRNA splicing in yeast. *Nucleic Acids Res.*, **35**, 4179–4185.
93. Müller,M., Weigand,J.E., Weichenrieder,O. and Suess,B. (2006) Thermodynamic characterization of an engineered tetracycline-binding riboswitch. *Nucleic Acids Res.*, **34**, 2607–2617.
94. Kim,D.S., Gusti,V., Dery,K.J. and Gaur,R.K. (2008) Ligand-induced sequestering of branchpoint sequence allows conditional control of splicing. *BMC Mol. Biol.*, **9**, 1–15.
95. Mandell, Douglas and Benett (2010) Principles and Practice Infectious Diseases.
96. Tulkens,P.M. (1999) Aminoglycosides : Nephrotoxicity. **43**, 1003–1012.
97. Melançon,P., Tapprich,W.E. and Brakier-Gingras,L. (1992) Single-base mutations at position 2661 of Escherichia coli 23S rRNA increase efficiency of translational proofreading. *J. Bacteriol.*, **174**, 7896–7901.
98. Vaara,M. (1992) Agents that increase the permeability of the outer membrane. *Microbiol. Rev.*, **56**, 395–411.
99. Bakker,P. (1992) Caused By the Insertion of Mistranslated Proteins Into the Cytoplasmic. *J. Gen. Microbiol.*, **138**, 551–561.
100. Andrew,W. (2013) Pharmaceutical manufacturing encyclopedia. *E1 Sci*.
101. WAKSMAN,S.A., LECHEVALIER,H.A. and HARRIS,D.A. (1949) Neomycin; production and antibiotic properties. *J. Clin. Invest.*, **28**, 934–939.
102. Sneader,W. (2005) Drug discovery: a history John Wiley & Sons.
103. Danziger,L.H. and Pendland,S.L. (1995) Bacterial resistance to beta-lactam antibiotics. *Am. J. Heal. Pharm.*, **52**, S3–S8.
104. DAVIES,J., GORINI,L. and DAVIS,B.D. (1965) Misreading of RNA codewords induced by aminoglycoside antibiotics. *Mol. Pharmacol.*, **1**, 93–106.
105. Tanaka,N. and IGUSA,S. (1968) EFFECTS OF VIOMYGIN AND POLYMYXIN B ON PROTEIN SYNTHESIS IN VITRO. *J. Antibiot. (Tokyo).*, **21**, 239–240.
106. Wank,H., Rogers,J., Davies,J. and Schroeder,R. (1994) Peptide antibiotics of the tuberactinomycin family as inhibitors of group I intron RNA splicing. *J. Mol. Biol.*, **236**, 1001–1010.
107. Schifano,J.M., Edifor,R., Sharp,J.D., Ouyang,M., Konkimalla,A., Husson,R.N. and Woychik,N.A. (2013) Mycobacterial toxin MazF-mt6 inhibits translation through cleavage of 23S rRNA at the ribosomal A site. *Proc. Natl. Acad. Sci.*, **110**, 8501–8506.
108. Infantino,V., Convertini,P., Cucci,L., Panaro,M.A., Di Noia,M.A., Calvello,R., Palmieri,F. and Iacobazzi,V. (2011) The mitochondrial citrate carrier: a new player in inflammation. *Biochem. J.*, **438**, 433–436.
109. Bauer,D.E., Hatzivassiliou,G., Zhao,F., Andreadis,C. and Thompson,C.B. (2005) ATP citrate lyase is an important component of cell growth and transformation. *Oncogene*, **24**, 6314.
110. Menga,A., Infantino,V., Iacobazzi,F., Convertini,P., Palmieri,F. and Iacobazzi,V. (2013) Insight into mechanism of in vitro insulin secretion increase induced by antipsychotic clozapine: role of

FOXA1 and mitochondrial citrate carrier. *Eur. Neuropsychopharmacol.*, **23**, 978–987.
111. Berg,J.M., Tymoczko,J.L. and Stryer,L. (2002) Glycolysis is an energy-conversion pathway in many organisms. *Biochem. 5th ed. New York WH Free.*
112. Folly,B.B. (2017) Fructose-1, 6-bisphosphate and its role on the flux-dependent regulation of metabolism.
113. Schinasi,L. and Leon,M.E. (2014) Non-Hodgkin lymphoma and occupational exposure to agricultural pesticide chemical groups and active ingredients: a systematic review and meta-analysis. *Int. J. Environ. Res. Public Health*, **11**, 4449–4527.
114. Mandal,M. and Breaker,R.R. (2004) Gene regulation by riboswitches. *Nat. Rev. Mol. Cell Biol.*, **5**, 451–63.
115. Nikolaus,N. and Strehlitz,B. (2014) DNA-Aptamers Binding Aminoglycoside Antibiotics. *Sensors*, **14**, 3737–3755.
116. Schneider,C. and Suess,B. (2016) Identification of RNA aptamers with riboswitching properties. *Methods*, **97**, 44–50.
117. Beeler,T., Bruce,K. and Dunn,T. (1997) Regulation of cellular Mg2+ by Saccharomyces cerevisiae. *Biochim. Biophys. Acta*, **1323**, 310–8.
118. Miyaguchi,H., Narita,H., Sakamoto,K. and Yokoyama,S. (1996) An antibiotic-binding motif of an RNA fragment derived from the A-site-related region of Escherichia coli 16S rRNA. *Nucleic Acids Res.*, **24**, 3700–3706.
119. Kozak,M. (1981) Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Res.*, **9**, 5233–5252.
120. Bar-Nun,S., Shneyour,Y. and Beckmann,J.S. (1983) G-418, an elongation inhibitor of 80 S ribosomes. *Biochim. Biophys. Acta (BBA)-Gene Struct. Expr.*, **741**, 123–127.
121. Kim,D.-S., Gusti,V., Pillai,S.G. and Gaur,R.K. (2005) An artificial riboswitch for controlling pre-mRNA splicing. *RNA*, **11**, 1667–1677.
122. Legiewicz,M., Lozupone,C., Knight,R. and Yarus,M. (2005) Size, constant sequences, and optimal selection. *Rna*, **11**, 1701–1709.
123. Rychlik,W. and Rhoads,R.E. (1989) A computer program for choosing optimal oligonudeotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res.*, **17**, 8543–8551.
124. Hüsken,D., Goodall,G., Blommers,M.J.J., Jahnke,W., Hall,J., Häner,R. and Moser,H.E. (1996) Creating RNA bulges: cleavage of RNA in RNA/DNA duplexes by metal ion catalysis. *Biochemistry*, **35**, 16591–16600.
125. Spill,F., Weinstein,Z.B., Irani Shemirani,A., Ho,N., Desai,D. and Zaman,M.H. (2016) Controlling uncertainty in aptamer selection. *Proc. Natl. Acad. Sci.*, **113**, 12076–12081.
126. Groher,F. and Suess,B. (2014) Synthetic riboswitches - A tool comes of age. *Biochim. Biophys. Acta*, **1839**, 964–973.
127. Vogel,M. and Suess,B. (2016) Label-free determination of the dissociation constant of small molecule-aptamer interaction by isothermal titration calorimetry. In *Nucleic Acid Aptamers*. Springer, pp. 113–125.
128. Duchardt-Ferner,E., Gottstein-Schmidtke,S.R., Weigand,J.E., Ohlenschläger,O., Wurm,J.P., Hammann,C., Suess,B. and Wöhnert,J. (2016) What a Difference an OH Makes: Conformational Dynamics as the Basis for the Ligand Specificity of the Neomycin-Sensing Riboswitch. *Angew. Chemie - Int. Ed.*, **55**, 1527–1530.
129. Woodside,M.T. and Block,S.M. (2008) Folding in Single Riboswitch Aptamers. *Science (80-.).*, **180**, 2006–2009.
130. Spiga,F.M., Maietta,P. and Guiducci,C. (2015) More DNA–Aptamers for Small Drugs: A Capture–SELEX Coupled with Surface Plasmon Resonance and High-Throughput Sequencing. *ACS Comb. Sci.*, **17**, 326–333.

131. Eaton,B.E., Gold,L. and Zichi,D.A. (1995) Let's get specific: the relationship between specificity and affinity. *Chem. Biol.*, **2**, 633–638.
132. Mandal,M. and Breaker,R.R. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat. Struct. Mol. Biol.*, **11**, 29–35.
133. Serganov,A. and Nudler,E. (2013) A Decade of Riboswitches. *Cell*, **152**, 17–24.
134. Wickiser,J.K., Cheah,M.T., Breaker,R.R. and Crothers,D.M. (2005) The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry*, **44**, 13404–13414.
135. Sonenberg,N. and Hinnebusch,A.G. (2009) Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. *Cell*, **136**, 731–745.
136. Schmidtke,S.R., Duchardt-Ferner,E., Weigand,J.E., Suess,B. and Wöhnert,J. (2010) NMR resonance assignments of an engineered neomycin-sensing riboswitch RNA bound to ribostamycin and tobramycin. *Biomol. NMR Assign.*, **4**, 115–118.
137. Rugbjerg,P., Genee,H.J., Jensen,K., Sarup-Lytzen,K. and Sommer,M.O.A. (2016) Molecular Buffers Permit Sensitivity Tuning and Inversion of Riboswitch Signals. *ACS Synth. Biol.*, **5**, 632–638.

9. Talks and poster presentation

MetaRNA symposium, Metabolism on the single cell level (2018), London, UK (Talk “Development of synthetic riboswitches using an optimized selection strategy to select structure-switching aptamers”)

Aptamers 2018 (2018), Oxford, UK (Poster price winner)

SFB 902 Symposium (2017), Friedriechsdorf, Germany (Talk: Development of synthetic riboswitch using RNA capture-SELEX, Presentation price winner)

Aptamers in Bordeaux (2017), Bordeaux, France (Poster)

33. Rabensteiner Kolleg (2017), Pottenstein, Germany (talk “Riboswitch engineering from capture-SELEX”)

Aptamers in Bordeaux (2016), Bordeaux, France (Poster)

10. Curriculum Vitae

Name	Adrien Boussebayle
Date of birth	31.01.1993 in Dax
Address	Lichtenbergstrasse 62 64289 Darmstadt Germany

Professional experience

09/2015 – present	**Institute for Biology | Synthetic genetic circuits | Technical University Darmstadt, Germany** Graduate student in the Lab of Prof. Dr. Beatrix Suess Doctoral thesis: *Development of RNA aptamers and synthetic riboswitch using Capture-SELEX* Degree: Dr. rer. nat.

Education

09/2013 – 08/2015	**Master of Science in Biochemistry interface biology, chemistry and physics** Master thesis: *Interaction characterization of the HIV integrase with its DNA aptamer and also with cofactors protein using Surface Plasmon Resonance* Degree: M. Sc.
09/2011 – 08/2013	**Bachelor of Science in Microbiology, cell and molecular biology** Degree: B. Sc.
09/2009 – 08/2011	**Technological University Degree, Biological engineering**

11. Ehrenwörtliche Erklärung

Ich erkläre hiermit ehrenwörtlich, dass ich die vorliegende Arbeit entsprechend den Regeln guter wissenschaftlicher Praxis selbstständig und ohne unzulässige Hilfe Dritter angefertigt habe.

Sämtliche aus fremden Quellen direkt oder indirekt übernommenen Gedanken sowie sämtliche von Anderen direkt oder indirekt übernommenen Daten, Techniken und Materialien sind als solche kenntlich gemacht. Die Arbeit wurde bisher bei keiner anderen Hochschule zu Prüfungszwecken eingereicht.

Darmstadt, den 15.09.2018

.....

(Adrien Boussebayle)